MICROSATELLITE MARKERS IN GENETIC IMPROVEMENT OF LIVESTOCK

A.P. USHA

Thesis submitted for the degree
of
Doctor of Philosophy

University of Edinburgh

1995

THE UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS

(Regulation 3.5.13)

Name of Candidate Areekkara Poduvattil Usha

Address Roslin Institute, Roslin, Midlothian EH25 9PS

Degree PhD *Date* 30.8.1995

Title of Thesis Microsatellite markers in genetic improvement of livestock

No. of words in the main text of Thesis approx. 45,000

Genome mapping is in the forefront of interest among both plant and animal breeders, enabling the relationship between genes, genome structure and function to be investigated in addition to identifying the location of genes. The prerequisite for a linkage map of the genome is the availability of a large number of highly polymorphic and informative marker loci which are evenly distributed throughout the genome. Microsatellite loci provide the unique class of markers which overcome many of the difficulties associated with the other marker types. The availability of a detailed genetic map of the bovine genome could enhance the genetic progress in cattle breeding programmes through the identification of loci affecting traits of economic importance. Other potential applications of genetic markers include their use in confirmation of parentage, individual identification, germplasm evaluation and identification of disease loci. In this thesis, microsatellite markers are investigated in three areas, a) parentage verification and individual identification b) study of phylogenetic relationship and c) mapping a lethal defect in Dexter cattle.

a) Five highly polymorphic microsatellite markers CYP21, DRB3, FSHB, ETH131 and HEL6 were evaluated for parentage verification using 275 animals belonging to 15 breeds of cattle. Some breeds were found not to be in Hardy Weinberg Equilibrium (HWE), the deviation being greatest in those breeds which had an excess of homozygotes. A new approach was developed for calculating the Probability of RAndom Sire Exclusion (PRASE) taking into account the deviation from HWE and linkage between markers, using observed genotype frequencies. Taken together, the linked markers, DRB3 and CYP21 gave a PRASE of 0.88 in all breeds with success ranging from 0.75-0.96. Including a third marker the PRASE was increased to 0.97, and with all five markers 0.99 or better was achieved for all the 15 breeds.

The possibility of simultaneous analysis or "multiplexing" the five markers was explored using an ABI 373 sequencer and fluorescently labelled primers. Co-amplification of three markers CYP21, DRB3 and FSHB and co-electrophoresis with ETH131 and HEL6 enabled the analysis of the five markers in a single gel lane. Application of these markers in paternity testing and individual identification is demonstrated and shown to be more robust than the commercialy available "Stock-marks" kit.

b) Allele frequencies of the five microsatellite loci were used to explore the genetic relationship between the cattle breeds. Genetic distances were calculated using Nei's, Cavalli's and Reynold's methods and phylogenetic trees were drawn based on these methods. Similarities between the trees were found but results were not always consistent. Microsatellite markers can be used for estimating genetic relationship in cattle breeds, however, five markers is the minimum for interpretable results and increasing the number used would give more consistent results.

c) The short legged diminutive Dexter cattle are a classical example of disproportionate dwarfism and it is suggested that they are the bovine equivalent of achondroplasia (ACH) in man. Three mutations in the transmembrane (TM) region of FGFR3 have been reported which cause ACH in man. FGFR3 was investigated as a candidate gene for dwarfism in Dexter cattle by linkage analysis, however, through inadequacy in pedigrees no conclusive linkage results were obtained. The TM region of FGFR3 was isolated from the "bull-dog" and unaffected Dexter cattle and sequenced enabling the three mutations causing ACH in man to be excluded in the Dexter phenotype. The bovine sequence of the TM region of FGFR3 was compared with that of human, mouse and chicken at the nucleotide and aminoacid level and showed 89%, 84.5% and 63.3% homology respectively.

The results presented in this thesis show that microsatellite loci are valuable tools in pedigree verification, individual identification, studies on population structure and the location of genes affecting bovine phenotype. The particular advantage of this type marker is the ease of analysis by PCR and that they have large number of alleles, reducing the number of loci to be used. Use of such polymorphic markers enable the identification of regions of the genome associated with production traits. However, ultimately as shown in the Dexter work, to understand the variation observed the responsible genes must be identified and characterised.

ACKNOWLEDGEMENTS

My sincere thanks to Dr.J.L.Williams, my supervisor for his advice and help in the conduct of this research work and his patience during the writing of this thesis.

My thanks are also to Prof.W.G.Hill, my University supervisor, for his encouragement and support and his keen interest in this work.

I am very grateful to Dr.Roger L. Spooner for his enthusiasm and constructive criticisms during this study and to Dr.Ramachandran for his help and guidance in the preparation of this manuscript.

My thanks and gratitude to Dr.S.P.Simpson, Dr.Chris Haley, Dr.Peter Keightely and Mr.Dave Waddington for helping me with the statistical analysis.

Special thanks to Dr.Douglas Lester, Dr.George Russell and Mrs.Barbara Urquhart for their help and valuable contribution for the conduct of this research.

I am grateful to my collegues in the laboratory Dr.Liz Glass, Bob Oliver, Janice Barr, Iain Mac Lean, Angela Gallagher, Susan Craigmile, Mrs. Olivia Small, Miss. Sarah Blott, David Brown, John Campbell, Suminder Sawhney, Parveen Goel and Victor Teres for providing me an enjoyable working atmosphere.

Thanks to Norman Russell and Elliot Armstrong for helping me in preparing the figures.

I am indebted to the Commonwealth Scholarship Commission in the UK for providing financial support to enable me to undertake this study.

Finally, my gratitude to my husband Satheesh, for his patience and encouragement during this work and who has spent more time than anyone in the preparation of this thesis and my son Jithin for his love and support.

Dedicated to my husband and son

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

RFLP	Restriction Fragment Length Polymorphism
PCR	Polymerase Chain Reaction
VNTR	Variable Number of Tandem Repeat
DNA	Deoxyribonucleic Acid
cDNA	Complementary DNA
mtDNA	Mitochondrial DNA
RNA	Ribo Nucleic Acid
tRNA	transfer RNA
rRNA	ribosomal RNA
mRNA	messenger RNA
PIC	Polymorphic Information Content
FISH	Fluorescent *in situ* hybridization
cM	Centimorgan
ul	microlitre
ug	microgram
mg	milligram
mM	millimolar
cm	centimeter
nm	nanometer
mCi	millicurie
Kb	kilo basepair
rpm	revolutions per miniute
MOET	Multiple Ovulation and Embryo Transfer
CEPH	Centre d'Etude du Polymorphisme Humaine
QTL	Quantitative Trait Loci
MAS	Marker Assisted Selection
GDD	Grand Daughter Design
FAO	Food and Agricultural Organisation
DUMPS	Deficiency of Uridine Monophosphate Synthase
DGGE	Denaturing Gradient Gel Electrophoresis
SSCP	Single Stranded Conformation Polymorphism
SSM	Single Stranded mispairing
PHI	Phosphohexose Isomerase

BLAD	Bovine Leukocyte Adhesion Deficiency
EMC	Enzyme Mismatch Cleavage
SDS	Sodium Dodecyl Sulphate
BSA	Bovine serum albumin
FMD	Foot and Mouth Disease
HWE	Hardy Weinberg Equilibrium
dATP	deoxy Adenosine Triphosphate
dCTP	deoxy Cytosine Triphosphate
dGTP	deoxy Guanosine Triphosphate
dTTP	deoxy Thymidine Triphosphate
dNTP	deoxy Nucleotide Triphosphate
ddATP	dideoxy Adenosine Triphosphate
ddCTP	dideoxy Cytosine Triphosphate
ddGTP	dideoxy Guanosine Triphosphate
ddTTP	dideoxy Thymidine Triphosphate
ddNTP	dideoxy Nucleotide Triphosphate
PRASE	Probability of RAndom Sire Exclusion
PMT	Photo Multiplier Tube
UPGMA	Unweighted Pair Group Method with Arithmetic means
NJ	Neighbor-joining
MP	Maximum Parsimony
ML	Maximum Likelihood
EDTA	Ethylene Diamine Tetraacetic Acid
DTT	Dithiothretiol
TEMED	N,N,N,N Tetramethylethylenediamine
APS	Ammonium Persulphate
PNK	Polynucleotide Kinase
BME	Betamercaptoethanol
ACH	Achondroplasia
FGF	Fibroblast Growth Factor
aFGF	acidic FGF
bFGF	basic FGF
FGFR	Fibroblast Growth Factor Receptor
IDUA	Iduronidase
TD	Thanatophoric dysplasia

PDEB	Beta Subunit of Phosphodiesterase
FSHB	Beta Subunit of Follicle Stimulating Hormone
HD	Huntingtons disease
HCH	Hypochondroplasia
JWS	Jackson-Weiss Syndrome
PFS	Pfeiffer Syndrome
CZS	Crouzon Syndrome

CHAPTER 1
GENERAL INTRODUCTION

1.1 The structure and variation of the bovine genome.

The bovine diploid genome consists of approximately 6,000 million base pairs of DNA, distributed over 29 pairs of acrocentric autosomes and a pair of sex chromosomes (Gallagher and Womack 1992). In addition, the mitochondrial genome is approximately 16500bp (Anderson et al. 1982). It is estimated that between 50,000 and 100,000 pairs of genes are encoded by the bovine genome, of which about 300 have been mapped to chromosomes (O'Brien et al. 1993a). The average bovine chromosome is approximately 100 million base pairs long containing 2,000 to 5,000 genes which consist of coding sequences, known as exons, interrupted by intervening sequences referred to as introns (Watson et al. 1992). Only about 5% of the genome is thought to actually code for protein, rRNA and tRNA (Caskey et al. 1989, Cunningham 1990). The remaining 95% of the DNA consists of various non-coding sequences (Jellinek and Schmid 1982). The complement of genes in the bovine genome is the blue print that determines the genetic makeup. McKusick (1980) regarded the genome as a part of the molecular anatomy of the cell. The molecular structure of the genome offers a powerful insight into the characteristics of the organism, because the genes are the blue print not only for its structure, but also for its functioning. Analysis of the genome is possible by studying variations in coding and non-coding sequences.

At the nucleotide level, there are variants at practically every locus. These variations constitute the so-called genetic polymorphism. To-date more than 1,000 polymorphic loci have been identified in the bovine genome (Barendse pers comm, Bishop et al. 1994, Georges et al. 1995). The nucleotide sequences of coding genes are fairly conserved and therefore show less polymorphism than non-coding introns and intergenic regions (O'Brien 1991). The non-coding regions of DNA are often repetitive sequences and are generally very polymorphic. It has been estimated that the genomes from two individuals vary by at least one nucleotide in every 300-1,000bp on average (Cooper et al. 1985). A large proportion of these sequence variants (30-40%) have been shown to occur as single base changes (Barker et al.

1984) or as variations in the number and type of repeat sequences. It is these polymorphisms that form the molecular basis of DNA genotyping.

1.2 DNA polymorphism.

Until 1988, analysis of DNA polymorphisms was essentially limited to the characterisation of restriction fragment length polymorphism (RFLP) (Georges et al. 1990a). Even though the existence of simple repeat motifs in eukaryotic genomes was known (Hamada et al. 1982), it was only with the introduction of the polymerase chain reaction (PCR) (Saiki et al. 1985) that analysis of these polymorphisms based on length variation in tandemly repeated DNA became popular. These polymorphisms are termed variable number of tandem repeats (VNTRs). Nearly 23% of the genome in higher animals consists of repetitive sequence of DNA identified as micro, mini, midi and macro satellites (Perret et al. 1990).

1.2.1 *Restriction Fragment Length Polymorphism.*

The first DNA polymorphisms to be widely used for genome characterisation and analysis were RFLPs (Grodzicker et al. 1974, Southern 1975). Restriction enzymes recognise specific sequences in the DNA and cause endonucleolytic cleavage where sites occur across the genome producing fragments of varying length. RFLPs are due to changes in the DNA sequence caused by loss or addition of a restriction site due to a base change or by deletion or insertion of sequences between restriction sites. The fragments are analysed in electrophoretic gels which separate them according to their length. The fragments are blotted onto a nitrocellulose membrane by Southern blotting (Southern 1975) and are identified by hybridisation methods with radiolabelled probes (Rigby et al. 1977). This is followed by autoradiography to detect the DNA fragments complementary to the probe. The techniques of Southern blotting, hybridisation and the use of probes are described in detail in chapter 2.

RFLPs have been used extensively to reveal polymorphism in DNA and to characterise populations of a variety of microbial, plant and animal species

(Andersen and Fairbanks 1990, Prince and Tanskley 1992, Karl et al. 1992, Megnegneau et al. 1993). These polymorphic markers were of great value in the contruction of early genetic maps (Botstein et al. 1980, Beckmann and Soller 1983, Donis-Keller et al. 1987, Georges et al. 1990a).

RFLPs occur frequently, follow mendelian inheritance, and are detected in co-dominant fashion. As coding sequences, they are conserved between species: probes derived from one species can usually be used to detect polymorphism in another species. Nevertheless, RFLPs have not been used extensively for livestock genome mapping, probably because it is expensive to develop the probes and their detection is a relatively laborious technique. RFLPs usually have a limited number of alleles with low heterozygosity and polymorphic information content (PIC). PIC is a measure of usefulness of a marker for linkage studies (Botstein et al. 1980) and is explained in chapter 3. Relatively large amounts of DNA are required for detecting RFLPs and several days to weeks are necessary to detect allellic patterns following hybridisation with radiolabelled probes (Southern blotting). Hence RFLPs are not suitable for routine typing. With the advent of PCR, the restriction digestion of PCR products and analysis on polyacrylamide gels has simplified the detection of RFLPs.

1.2.2 *Variable Number of Tandem Repeats (VNTRs).*

VNTRs are sequences that occur repeatedly throughout the genome and are highly polymorphic, due to variation in the number of repeats at a particular locus in different individuals (Jeffreys et al. 1985a, Nakamura et al. 1987). Polymorphic information content is generally higher for VNTRs than for RFLPs. VNTR loci are divided into two classes: minisatellites and microsatellites. Whilst both mini and microsatellites are VNTR loci, the term VNTR is more commonly associated with the former, and the term short tandem repeat or STR loci for microsatellites. The difference between these markers are in the length and size of the individual repeat units: 15-60bp for minisatellites (Jeffreys et al. 1985a) and 2-5bp for microsatellites (Litt and Luty 1989). These markers seem to differ in their genomic distribution. In cattle, minisatellites are clustered in the telomeres of chromosomes (Royle et al.

1988, Nakamura et al. 1988), whereas microsatellites appear to be distributed throughout the genome (Weber and May 1989). Another major difference between these loci is the method of genotyping. Minisatellites are analysed by Southern blotting in the same manner as for RFLPs, but microsatellites are characterised by PCR amplification and polyacrylamide gel electrophoresis.

1.2.2.1 *Minisatellites.*

Following the first description of a highly polymorphic locus in human DNA (Wyman and White 1980), hypervariable regions were identified in human insulin (Bell et al. 1982), globin (Higgs et al. 1981) and apolipoprotein B genes (Knott et al. 1986). Analysis of the nucleotide sequence showed that the structural basis for the polymorphism in these regions was the presence of a variable number of short tandem repeats that could be detected as RFLPs with any restriction enzyme whose recognition sequence was not present in the VNTR. Hence the VNTR is a special type of RFLP. The partial homology that was noted between the repeats at different loci led investigators to postulate that something inherent in the sequence might allow tandem expansion at the independent loci (Proudfoot et al. 1982, Goodbourn et al. 1983). Jeffreys et al. (1985a) described two such simple repeat probes which detect multiallelic loci dispersed at many locations throughout the genome. They used the 33 basepair myoglobin repeat in hybridization experiments to screen a human genomic library and demonstrated that this sequence cross-hybridized with several related sequences and defined the first identified human family of minisatellite loci. The banding pattern shown by different individuals varied, indicating that the probes were hybridizing to VNTR regions. The banding profile was individual specific and was called a DNA or genetic fingerprint. In addition to the probe from the repeat in the myoglobin gene, other naturally occuring probes - M13 and alpha globin (Vassart et al. 1987, Fowler et al. 1988) and synthetic tandem repeat probes (Ali et al. 1986, Vergnaud 1989) have been successfully used to detect multiple hypervariable loci in human DNA.

The minisatellite probes detecting multiple VNTR regions are not suited for linkage studies because of the difficulty in attributing hybridization bands to specific loci (Jeffreys and Wilson 1985). This led to the development of single

locus-specific VNTR probes (Wyman and White 1980, Nakamura et al. 1987) by direct cloning of the high molecular weight bands of multilocus DNA fingerprints (Wong et al. 1986,1987). Good success has been reported with these locus-specific VNTR probes because the DNA fingerprints uncovered by these probes differ from one another and from those seen by the Jeffreys and M13 probes.

The presence of a large number of alleles at minisatellite loci renders them as valuable tools in the establishment of family relationships (Jeffreys and Wilson 1985, Jeffreys et al. 1991a) because different individuals will have distinct pattern of fragments (Armour and Jeffreys 1992). Because of high heterozygosity index, approaching 90% (Wong et al. 1987), these loci have proven to be very powerful tools in many applications such as forensic medicine (Gill et al. 1985, 1987, Jeffreys et al. 1985b), paternity testing (Jeffreys and Wilson 1985), linkage analysis and mapping studies (Jeffreys et al. 1986, Perret et al. 1990), and the determination of twin zygosity (Hill and Jeffreys 1985). In addition to the uses in human genetic analysis, multi-locus DNA fingerprints also have an important application in dogs, cats, birds, cattle etc (Jeffreys and Morton 1987, Burke and Bruford 1987, Georges et al. 1990a).

Minisatellite probes in man were shown to hybridize to hypervariable loci in domestic animals (Jeffreys and Morton 1987, Georges et al. 1988b, 1990a). Minisatellites have been suggested to have an important role in animal improvement (Soller 1990) in various aspects of individual identification (Buitkamp et al. 1991), parentage identification (Georges et al. 1988a, Haberfeld et al. 1993), identification of genetic disorders (Georges et al. 1990a) and linkage analysis (Georges et al. 1988b). However, the use of minisatellites in linkage analysis in cattle is limited because of the preferential telomeric location of these loci (Royle et al. 1988, Nakamura et al. 1988). PCR analysis of length variation at minisatellite loci is possible but suffers from the disadvantage that many alleles at highly polymorphic loci are too large (20-25kb) to amplify efficiently (Jeffreys et al. 1988). Therefore detection is expensive and relatively laborious relying on Southern blotting. In addition, the comparison of results between laboratories is difficult, because size estimates of bands in the DNA fingerprints are not always reproducible (Jefferys et

al. 1991b).

1.2.2.2 *Microsatellites.*

The dinucleotide repeat sequence (dA-dC)n (dG-dT)n represents one of the most abundant families of interspersed repetitive DNA in eukaryotic genomes (Miesfield et al. 1981). Variation of a dinucleotide repeat unit in a malarial parasite antigen (Kemp et al. 1987, Weber 1988) prompted the study of other (dA-dC)n (dG-dT)n sequences, and it was found that the repeats are polymorphic and informative (Weber and May 1989, Smeets et al. 1989). The existence of short tandem repeats such as d(A)n, d(T)n, (dT-dC)n, (dA-dG)n and (dTTA)n, d(AAT)n have also been described by several investigators (Tautz 1989, Litt and Luty 1989, Economou et al. 1990, Zuliani and Hobbs 1990, Stallings et al. 1991). It is clear that these loci can be used as genetic markers, based on the variability in length of the short tandem repeat sequences or microsatellites (Weber and May 1989, Litt and Luty 1989).

Weber (1990a) examined over 100 human poly CA repeats and classified the microsatellite regions into three categories: perfect repeat sequences without interruption in the dinucleotide repeat (65% of total), imperfect repeat sequences with one or more interruptions in the repeat (25%) and compound repeat sequences with adjacent tandem simple repeats of different sequences (10%). Most of the blocks of human microsatellites are relatively short in length. The longest uninterrupted repeat found to date contains 37 CA repeats (74bp), but a vast majority have 25 or fewer repeats (Hamada et al. 1984a, Gross and Garrard 1986, Weber 1990a).

Microsatellites sequences have been found in all eukaryotes examined but not in prokaryotes (Hamada et al. 1982, Gross and Garrad 1986). On average, microsatellites occur in a haploid mammalian genome as 5-10 X 10^4 individual islets (Litt and Luty 1989, Beckman and Soller 1990, Stallings et al. 1991). If these islets are randomly distributed over the 3 X 10^9bp of the bovine genome, one would expect a microsatellite sequence in every 50-100kb on average (Beckmann and Soller 1990, Weissenbach 1993).

In 1989, three groups of workers simultaneously discovered that microsatellites are highly polymorphic in eukaryotic genomes (Litt and Luty 1989,

Tautz 1989, Weber and May 1989). The microsatellies show site-specific length variation and hence, on average, a PIC of 0.61, which makes them versatile markers for genome mapping. PIC is calculated from the allele frequencies in the population and is found to be positively correlated with the length of repeat (Weber 1990a, Hazan et al. 1992) ie for longer repeats more alleles are expected.

Microsatellite sequences can be amplified *in vitro* using PCR (Weber 1990b) or can be analysed in a similar fashion to minisatellite probes (Haberfeld et al. 1991, Trommelen et al. 1993). By PCR, these polymorphisms are revealed by amplification of genomic DNA using primers flanking the chosen repeat sequence and specific for a given locus (Cornall et al. 1991, Kemp and Teale 1991). The alleles (which typically differ in length by 2bp) can be differentiated by fractionating the amplified DNA product on a high resolution gel. Automated DNA typing methods have been developed which simplify detection and analysis of microsatellite polymorphism (Ziegle et al 1992).

Mutation rates in microsatellites are calculated to be approximately 10^{-4} per generation in man and mouse (Weber 1990a, Dallas 1992). Intra-allelic slipped strand mispairing is currently the most favoured mechanism for mutation in simple sequence tandem repeats (Levinson and Gutman 1987). This mechanism of mispairing of tandem repeats is explained in chapter 3. Unequal recombination between homologous chromatids during meiosis is also a possible cause of mutation in microsatellites (Baron et al. 1992).

The function of the $(CA)_n$ blocks is unknown, but there are reports which indicate that these sequences could enhance the expression of an adjacent gene especially when the repeat sequence is close to the promoter sequence (Hamada et al. 1984b, Glaser et al. 1990). It is also proposed that these repeat sequences serve as hotspots for recombination (Slightom et al. 1980). The absence of $(CA)_n$ repeats in prokaryotic genomes suggest that they might have a role in packaging and condensing DNA in eukaryotes. Moreover, the expansion of trinucleotide repeats as the cause of dominantly inherited neurodegenerative disorders (Mahadevan et al. 1992, La Spada et al. 1991, Richards and Sutherland 1992) assigns a pathogenic role to some microsatellites (Stallings et al. 1994).

Even though no specific function has been assigned to microsatellite sequences, they have proved to be efficient markers for the mapping of economic trait loci (Georges et al. 1994) and disease genes in animals (Holmes 1994). The use of these markers for parentage determination (Glowatzki Mullis et al. 1995, Usha et al. 1995) and evolutionary studies (MacHugh et al. 1994) has also been suggested. Thus these markers have proved to be of great use for studies in population and ecological genetics, gene mapping and medical genetics, and are currently the favoured markers in human and animal genetic research.

1.3 Genome mapping.

Mapping the genome enables one to understand the relationship between gene/genome structure and function and to identify the location of genes. The analysis of the genome of livestock species will eventually lead to the understanding of the genetic control of their various physiological traits. Genome mapping is the first requirement for a detailed genome functional analysis (Barendse et al. 1994, Bishop et al. 1994, Georges et al. 1995).

The prerequisite for genome mapping is the availability of a large number of highly polymorphic marker loci which are evenly distributed along the genome. The inheritance of these markers is followed in families for the construction of genetic maps. Mapping proceeds by determining the relationship between markers based on segregation analysis. The efficiency with which the map can be used depends on the value of the markers (number of alleles) and the resolution of the map (the distance between the markers). A distance of 20cM between markers is sufficient to detect markers linked to QTL in families in which the trait segregates (Botstein et al. 1980). The bovine genome is estimated to be 2800 cM, and therefore to cover the genome at 20 cM resolution requires about 140 evenly spaced markers.

The construction of animal genome maps has two distinct goals. Firstly, they are the tools for genetic analysis and manipulation. Primary interest in the bovine genome map, however lies in its potential use as a means of mapping loci affecting production characters and disease susceptibility or resistance and for manipulating these loci in breeding programmes (Smith and Simpson 1986; Beckmann and Soller

1987). Secondly, genetic maps help to dissect the evolution of the mammalian genome organisation and the relationship between taxonomic groups. Thus mapping the bovine genome is also of significant interest as a study in comparative chromosome organisation.

1.3.1 Methodology of bovine genome mapping.

Genome mapping can be approached in two different but complementary ways: physical and genetic (linkage). Different methodologies used for bovine genome mapping are reviewed by Fries et al. (1989) and Fries (1993). The construction of physical maps is based on the methods that make use of chromosome transfers between mammalian somatic cells from different species (somatic cell hybridization), and methods that involve the direct assignment of genes by molecular hybridization of specific DNA probes hybridised to metaphase chromosomes (*in situ* hybridization). There are also comparative investigations that allow provisional mapping based on the evolutionarily-conserved location of the genes. Genetic maps, on the other hand, are constructed by studying the segregation of loci in families (linkage analysis) and provide information about the distance of genes as a function of the meiotic crossovers occuring along the chromosomes.

1.3.2 *Physical mapping.*

1.3.2.1 *Somatic cell hybridization.*

The assignment of genes to particular synteny groups represents the first level of physical mapping. The first reported attempt to assign a gene to a particular autosome in cattle was by somatic cell hybridization (Dain et al. 1984). The technique involves creating hybrid cells by fusing cells from the species of interest with an immortalized cell line, usually of rodent origin. Chromosomes from the donor cells tend to be lost in the hybrid cells, while the rodent component is generally stable. Panels of hybrid cell lines are generated and mapping is achieved by correlating the presence or absence of the genes of interest with others assigned to particular cells in the panel. The analysis of gene products (by electrophoresis) or genes through Southern hybridization (Gallagher et al. 1991), or PCR

amplification (Dionne et al. 1990) in these hybrid lines allows the identification of groups of genes that are syntenic. The genes which lie on the same chromosome are found in the same cells and are said to be syntenic. In cattle, synteny groups for all chromosomes are defined (Fries et al. 1993) and are designated as U1 to U24 and U26 to U29 (the 'U' standing for unassigned to chromosomes). Earlier studies with somatic cell hybrid lines allowed the assignment of genes to synteny groups and the examination of conservation of synteny between man, mouse and cattle (Womack and Moll 1986, Womack et al. 1989).

1.3.2.2 *In situ hybridization.*

In situ hybridization of karyotypes provides another means for the physical mapping of genes or DNA fragments to specific regions in the chromosomes (Fries et al. 1989). It involves the hybridization of radio or fluorescently labelled DNA probes (labelled with tritium and biotin-11-dUTP respectively) to fixed metaphase chromosomes (Zneimer and Womack 1989, Fries et al. 1991). The probe hybridises to the complementery sequences on individual chromosomes and the hybridization is detected by silver grains on the autoradiograph or as fluorescence as in the case of fluorescent *in situ* hybridization (FISH). *In situ* hybridization enables unassigned synteny groups detected by somatic cell hybrids to be assigned to different chromosomes of the bovine genome (Fries et al. 1988, Barendse et al. 1994).

1.3.3 ***Genetic mapping.***

Genetic maps date back to the work of Morgan and co-workers in the early part of this century. They found that in *drosophila* certain characteristics (body colour and wing size, eye colour and eye shape) were inherited together rather than following independent segregation. These characterestics were described as being "linked". The degree of linkage was calculated allowing the location (mapping) of the relative position of the genes on the chromosomes. By 1915, 85 characterestics were mapped in *drosophila* into four linkage groups, one on each of four pairs of chromosomes (Gardner et al. 1991). Since then linkage analysis has been conducted in all mapping studies.

The genes that lie closer to each other in a linear array along the

chromosomes have a less opportunity for recombination. Genes that recombine once in every 100 meiotic events are said to be one centimorgan (1cM) apart. This relationship between the linear proximity of genetic loci and the recombinational frequency between them provides the basis for linkage mapping (Morton 1955, Botstein et al. 1980, White et al. 1985). Linkage maps are constructed by measurement of meiotic recombination frequency between loci in informative families. If the alleles at two loci co-segregate in the progeny the two loci are likely to be linked. The greater the distance between the two loci, the more are the chances for recombination events to be observed. The lod score test (logarithm of the odds test: Morton 1955, Lathrop et al. 1984) is now widely used to determine whether linkage can be accepted or must be rejected for a set of segregation data. The lod score test is explained in section 1.4.3.2 of this chapter. The map distance is measured in crossover units, or centimorgans, and can be estimated on the basis of frequency of recombination using an appropriate mapping function. As recombination is more frequent in females than in males over particular sections of the genome (Dunn and Bennet 1967), the female linkage map is generally longer than the male linkage map. Correlation between the actual physical length and the linkage map length varies across the genome, a figure of 1cM as an average equal to approximately one million base pairs is used as a rule of thumb. Moreover, the figure is species-specific, because different organisms differ in frequency of recombination per unit of DNA. For example, in terms of map units, the yeast genome is nearly five times longer than that of maize, but its DNA content is nearly 500 times less (Fincham 1983).

In order to construct a genetic map there is a requirement for informative families or pedigrees of appropriate sizes and polymorphic genetic markers. The ideal mapping family consists of large full sibships. However, the formation of large full sib families is slow in cattle because they are monoparous and have long generation interval. Nevertheless, artificial insemination and multiple ovulation and embryo transfer (MOET) have overcome these problems to a certain extent.

1.3.4 Reference families.

One of the most efficient means of developing a genetic map is by the collaborative analysis of mapping pedigrees shared on an international basis. Because many markers are involved in mapping experiments, distinct families are important in which alleles of various loci are segregating. For this reason, "reference families" were established by human geneticists (White and Lalouel 1988). "Reference families" are those used as a common resource by research groups for mapping genes or genetic markers by linkage analysis. The family structures should therefore be efficient for mapping a wide variety of genes or markers from both high and low polymorphic systems. "Resource families" are those in which specific genes or traits of biological or economic importance, controlled by a number of genes are segregating or likely to be segregating. A typical example of a "reference family" data set is found in the Centre d'Etude du Polymorphisme Humaine (CEPH) in Paris promoting the development of the human genetic map by making DNA samples available from 40 families over many generations (Dausset et al. 1990). A reference material is a collection of samples from members of informative families. The samples (DNA, tissues, cell lines) are stored and are available for research. This type of informative material has been described to study marker gene effects on the values of quantitative traits in crossbred generations of mice (Kluge and Geldermann 1982, Bulfield et al. 1984).

Hetzel (1991) emphasized the importance of a collaborative effort between laboratories to maintain informative families for cattle. For construction of the genetic linkage maps in cattle an initial international reference panel of full sib families provided from CSIRO, Australia and Texas A & M University have been used since 1989 and 1991 repectively. In 1993, two large full sibship families were included from ILRAD, Kenya. The international reference panel was enlarged subsequently with the inclusion of families from Europe. The parental genotypes are *Bos taurus* and *Bos indicus* breed crosses. 295 individuals are included in the current international bovine reference panel.

1.3.5 Genetic markers.

Genetic markers have been classified as type I anchor loci and type II loci (O'Brien 1991). Type I anchor loci are coding sequences and are conserved across species. The type I marker map allows one to identify the boundaries of conserved synteny, and also provide candidate genes once a trait locus has been assigned to a chromosome. Even though the type I loci may be important for their effects on genetic traits, these markers are not immediately useful for mapping, because of their limited polymorphism. Type II loci are non-coding anonymous DNA sequences which are species-specific and are used to map trait loci. These include the minisatellite and the microsatellite loci. The close evolutionary relationships between bovidae (cattle, sheep and goats) enable detection of the conservation of synteny with these polymorphic type II loci. The genetic maps developed through the collaborative analysis of reference families feature both type I and type II markers.

1.3.6 Comparative gene mapping.

The map of the human genome is more extensive than those of other species, and contains a considerable number of genes whose products and functions are known. Comparison of genomic maps between species reveals that large regions are conserved in terms of the genes present and their order ie. conservation of synteny (Nadeau and Taylor 1984, Womack and Moll 1986, Womack 1987). This suggests that the knowledge obtained through the study of one species can be used to predict the location of loci in another and thus speed up the mapping procedure (Pepin et al. 1995).

Womack and Moll (1986) reported a high degree of conservation of synteny between cattle and human chromosomes. Alignment of the regions conserved between man and cattle will enable information from the human maps to be used in studies with cattle, eg: the homologues of genes found on the short arm of human chromosome 11 (11p) and on the mouse chromosomes 2 and 7 are found on chromosome 15 in cattle (Fries et al. 1993). Other examples of the high degree of conservation are human chromosomes 9 and 12 which are highly conserved in cattle

to form chromosomes 8 and 5 (O'Brien 1991, Barendse et al. 1994). In contrast, the same genes are found on three mouse chromosomes (Threadgill and Womack 1990a,b). The comparison of human, mouse and cattle genomes suggests that the conservation of synteny between cattle and man is higher than that between cattle and the mouse (Fries et al. 1993).

A special aspect of comparative mapping is the high degree of karyotype conservation in bovidae (Gallagher and Womack 1992). Comparative gene mapping and cytogenetics in mammalian groups can also provide information on the mechanism of chromosomal rearrangements during evolution (Threadgill et al. 1991) and thus serve as a tool for the understanding of the evolution of the mammalian genome. Mapping several genes in cattle, sheep and goat has revealed that homologous genes usually map to karyotypically homologous chromosome regions (Hediger et al. 1991). The comparative mapping approach has made considerable progress in the development of the bovine genome map. Because of the conservation of synteny between bovine and ovine sequences, primers used to detect microsatellite polymorphisms in cattle are often directly useful to reveal polymorphisms in sheep and vice versa (Moore et al. 1991).

1.3.7 Progress in the bovine genome map.

The bovine genome is the most extensively studied farm animal genome and ranks fourth after the genome maps of man, mouse and rat. The genome map of cattle has developed fairly rapidly in the recent years using a combination of techniques such as somatic cell hybrid panels, *in situ* hybridisation, linkage analysis and comparative mapping.

Until 1989, 26 syntenic groups of autosomal bovine genes were identified and 9 syntenic groups were assigned to specific autosomes (Dain et al. 1984, Womack 1990). Approximately 200 genes were assigned to 29 syntenic groups by 1991 and by 1993, 300 genes were assigned to syntenic groups. Genes, and consequently the syntenic groups, were assigned to 17 chromosomes including the X and Y chromosomes.

Fries et al. (1993) reported a total of 350 bovine loci assigned to a

chromosome, synteny group or a linkage group. Of these, 248 contained coding genes and 93 were anonymous loci. For these loci 28 autosomal (U1-U24, U26-U29), 2 sex chromosome specific synteny groups (X,Y) and 28 linkage groups (L1-L17, L19-L22, L26-L32) were identified. Of the polymorphisms determined, 84 were VNTR loci, 24 were minisatellites and 60 microsatellites. A primary genetic map was reported by Barendse et al. (1994) containing 202 markers, of which 144 were microsatellites; 114 of these were assigned to a physical map covering 90% of the expected length of the cattle genome. Simultaneously, the map by Bishop et al. (1994) reported a total of 313 polymorphic markers of which 172 were microsatellites and 16 protein-coding genes were assigned to individual chromosomes. The current genetic linkage map consists of about 700 DNA markers, the majority of which are microsatellite loci. 136 DNA polymorphisms at the coding sequences, have been mapped. Also 123 markers have been physically mapped, most of which are genes. All 30 chromosomes have been assigned to syntenic and linkage groups and are identified by *in situ* hybridisation of markers.

1.4 Use of genetic markers in animal improvement.

The genetic map of the bovine genome being very well covered with highly polymorphic genetic markers (Bishop et al. 1994, Barendse et al. 1994) could, in principle, enhance the genetic progress in cattle breeding programmes (Soller and Beckmann 1983, Hallermann et al. 1986, Kashi et al. 1986, Womack 1987). The potential applications of genetic markers include their use in confirmation of parentage of individuals and line identification purposes (Soller and Beckmann 1983, Kashi et al. 1990a, Trommelen et al. 1993), identification of genetic loci affecting traits of economic importance (Beckmann and Soller 1987, Womack 1993), germplasm evaluation (Soller and Beckmann 1983) and identification of the disease loci (White and Lalouel 1988). The principle of using genetic markers in breeding programmes is that they allow the inheritance of whole segments of chromosomes to be traced from parents to offspring. The use of markers for confirmation of parentage and germplasm evaluation are discussed in chapter 3 & 5 respectively.

The polymorphic type II markers have by themselves no effect on

performance, but they mark chromosome segments containing genes which may affect performance. If a particular marker is associated with a gene affecting performance trait, it allows the progeny to be selected on the basis of inheritance of these markers from their parents. The genes that are followed using markers may be those causing qualitative differences between animals (eg: genes for coat colour, horns etc) or may be those contributing to quantitative variation between animals for a performance trait such as milk production.

In animal breeding programmes three major applications for genetic markers have been suggested: marker assisted selection, marker assisted introgression and identification of genes associated with diseases (Georges 1991).

1.4.1 Marker Assisted Selection.

With the advent of artificial insemination, the impact of superior sires on the genetic progress of breeds is potentially enormous. Widespread use of a given sire, however, is only justified when its breeding value has been estimated with sufficient reliability. Until now, this has required progeny testing. Young sires, resulting from planned matings of sires and dams with highest breeding value, are tested on the basis of the milking performances of 50-100 of their daughters. This is a time-consuming and expensive procedure. Approximately 5-6 years elapse between the time of selection of parents of a candidate sire and the estimation of its breeding value from its daughters records. Although this procedure has proven to be very effective in improving the genetic merit of the herd, it is necessary to develop faster and cheaper methods to predict the genetic value of an animal.

Genetic improvement of animal populations is limited by the fact that most traits of economic importance are polygenic in nature and are influenced by a variety of environmental and genetic factors (Pirchner 1983). Consequently, it is generally not possible to determine the genotype of any particular individual with respect to economic traits by examination of phenotype alone. One possibility to improve the genetic gain is by identifying the quantitative trait loci (QTL) (Geldermann 1975) contributing to the genetic variants of production traits. Recent advances in theoretical quantitative genetics suggest that genetic gain can be achieved using

genetic markers that are closely linked to genes that affect quantitative traits (Geldermann 1976, Soller and Beckmann 1983). This information could then be used to select animals based on their genotype at the QTL in a procedure called marker assisted selection (MAS) (Soller and Beckmann 1982, Smith and Simpson 1986).

A potential application of a QTL map will be the marker assisted selection of young bulls prior to progeny testing (Geldermann 1975, Soller 1978, Smith and Simpson 1986, Kashi et al. 1990b), by identifying those which inherit the alleles for marker loci known to be linked to a favourable allele at the QTL. It has been shown that MAS of candidate bulls based on the use of single diallelic marker linked to a QTL only has a negligible effect on the rate of genetic progress, and is increased by 20-30% by using haplotype of diallelic and polyallelic markers (Kashi et al. 1990b) because a greater proportion of sires will be heterozygous at the marker locus and a greater proportion of daughters will be informative for the marker/QTL linkage phase. Thus the polymorphic markers associated with a favourable allele for a trait helps breeders to select bulls for progeny testing at a very young age (Soller and Beckmann 1982, 1985). Marker-assisted selection increases the scope for selecting QTL and making selection decisions before the trait is manifested eg: milk yield and longevity (Soller 1990).

1.4.1.1 *Genetic markers and animal improvement.*

Attempts to find associations between blood groups and production traits started as early as 1950, ten years after the first report of cattle blood groups by R.Irwin (Mc Clure 1952, Nair 1957). The results obtained were not conclusive, because of the lack of statistical and biological significance with too few animals and too few markers. These loci cover only a small part of the genome and it is extremely unlikely that these markers will be close to a QTL. Even though many studies have been conducted using blood groups and biochemical markers associated with milk production traits (Niemann-Sorrensen and Robertson 1961, Rendel 1961, Geldermann et al. 1985, Gonyon et al. 1987, Haenlein et al. 1987) and carcass characteristics in cattle (Beever et al. 1990), the effects found are not large enough to be useful in selecting high yielding stock (Hines 1990). Later, DNA-based

markers such as restriction fragment length polymorphisms (Lander and Botstein 1989, Paterson et al. 1988), minisatellites (Georges et al. 1990a) and microsatellites (Jacob and Hwang 1992) have been used as genetic markers to localise quantitative and qualitative trait loci.

Marker mapping has been used to search for markers for monogenic traits which are associated with quantitative traits. The association results from either the pleiotropic effect of a single gene on both monogenic and the quantitative trait, or from genetic linkage between the gene underlying the monogenic trait and a QTL. Examples of such an association include hyperkalemic periodic paralysis and muscularity in quarter horses (Rudolph et al. 1992), malignant hyperthermia and the porcine stress syndrome (pale soft exudative meat syndrome) in pigs (Mac Lennan and Philips 1992), progressive degenerative myeloencephalopathy or Weaver Disease in Brown Swiss cattle (Hoeschele and Meinart 1990) and increased milk production in cattle (Georges et al. 1993a). The genetic map of cattle is now of sufficient resolution that it can be used to identify markers associated with QTL. Many laboratories are currently involved in QTL experiments based on the bovine genome map and have found associations with several economic traits (Ron et al. 1994, Georges et al. 1995).

Even though the genomic map will provide the tool to locate loci controlling phenotypic variation, in order to apply the map, it is necessary to have informative groups of animals segregating in the trait of interest. The number of animals required depends on the number of genes controlling the traits and the phenotypic variation. Markers for a monogenic trait could be identified using as few as 50-100 individuals in half-sib family groups. However, most of the quantitative traits selected for livestock, for example, milk yield are polygenic in nature, ie. the individual phenotype reflects the action of several genes. Because the contribution of individual QTL to the overall phenotypic variance may be low, their mapping will require very large number of informative individuals.

Two methods, selective genotyping and replicated progenies are the possibile ways of reducing the number of offspring scored against each marker to identify quantitative trait loci. Selective genotyping is based on the observation by Stuber

et al. (1980) that selection for quantitative traits changed marker allele frequencies in segregating populations. This method can be applied to dairy cattle by selecting only high and low producing daughters of each sire for marker scoring with respect to one or two of the most important traits. Even though selective genotyping would increase the power markedly with respect to the selected trait, the method cannot be applied to two or more independent traits simultaneously, unless the proportion selected is very small, and this would require the entire population for the markers to be scored (Lebowitz et al. 1987).

The second method is of progeny testing the offspring in order to decrease the error variance of quantitative trait evaluation (Soller and Beckmann 1990). This can be readily applied to dairy cattle data, because a small number of elite sires are used for breeding in routine progeny tests. Two experimental designs, daughter and grand daughter design, are used. In the first, the merit of the sire is evaluated by the performance of their daughters. This design may not be appropriate for low heritability traits for which the grand daughter design (GDD) is applied (Weller and Fernando 1991). This method is described by Weller et al. (1990) and uses three generation families to reduce environmental variance. The sons of a popular sire are genotyped, and the genetic merit of the sons is determined by the performance of their daughters. The genetic merit of the sons is then compared with the markers segregating from their sire to associate markers with the traits. In a grand daughter design experiment, Ron et al. (1994) noted a significant effect of the marker D21S4 on milk production and protein content in one sire using Israeli Holstein families. In another study, Georges et al. (1995) used 14 grand sires with a total of 1518 sons to examine the association of 159 markers with protein and fat contents in milk in American Holstein cattle. In this study, five chromosomes (1, 6, 9, 10 and 20) were identified with strong evidence for the presence of a QTL controlling milk production.

1.4.1.2 *Parentage verification*

Prior to the 1980's, parentage determination was based on blood group data and protein polymorphisms. The accuracy of the blood tests is reported to be high; but their usefulness was limited due to the finite number of blood group systems and

the low level of polymorphisms in serum proteins. Detailed description of these markers is given in chapter 3. The introduction of DNA based technologies have lead to the identification of markers far more efficient than erythrocyte antigens and serum proteins.

Although RFLPs and minisatellites are better markers for parentage and individual identifications, the technical difficulties associated with their use have rendered them less suitable for routine testing (Allen et al. 1990, Wenk et al. 1992). With microsatellites, high levels of polymorphism and the relative ease of processing the test samples make them highly suitable in forensic medicine. Considerable progress has been made in automated typing using fluorescent labelled microsatellite markers (Ziegle et al. 1992). Early in 1993, in the International comparison test panels of different microsatellite markers were compared for the first time between laboratories.

Different well established statistical procedures are available for calculating the probability of exclusion of incorrect sires in parentage tests. In the present study, the usefulness of five microsatellite loci has been examined for parentage verification in a significant number of cattle belonging to 15 breeds taking into consideration the deviations from HWE and linkage equilibrium and using appropriate statistical methods relevant to small sample size.

1.4.2 Marker assisted introgression.

No single commercial line or breed is likely to contain all of the best alleles for all traits of economic importance. If alleles of superior value for one or for a few traits can be identified in breeds which are inferior in economic merit, an efficient crossing programme using breeds with superior economic value might increase overall performance substantially. This process of moving desirable genes from one breed can be more efficient using marker assisted introgression and is achieved by selective backcrossing between these breeds (Soller and Beckmann 1982, 1990).

For introgression, genetic markers are used for the gene to be introgressed and to select for (or against) a particular background genotype. A desired gene and

genetic markers associated with it in an inferior breed is located and by introgression the favourable alleles are fixed in a commercial population which is superior in economic value. This is done over a number of generations of backcrossing of a population which carries the allele to be introgressed to a recipient population followed by an *inter se* cross to make the desired allele homozygous (Visscher and Haley 1995). The selection for background genotype is done in two stages.

Introgression has been less commonly applied in domestic animals than in plant species, because of the limited fertility, generation interval and the greater expense of each individual. The Food and Agricultural Organisation (FAO) has explored the potential of marker-assisted introgression as a means of introducing the gene for trypanotolerance from the tolerant N'dama breed of West Africa to other susceptible tropical breeds (Beckmann and Soller 1987). Marker assisted introgression of specific mapped QTL can be an effective means of upgrading native breeds while retaining adaptation to local conditions (Beckmann and Soller 1983, Weller and Fernando 1991). The efficiency of marker assisted introgression programmes depend on the frequency of the introgressed alleles in the final population.

1.4.3 Genetic markers and diseases.

The application of molecular biology has revolutionized our understanding of the genetic basis of disease susceptibility in both man and animals. Before the advent of DNA technology, geneticists used blood group and protein polymorphisms, visible cytogenetic abberations (deletions, duplications, translocations) and other phenotypic markers to map hereditary diseases. For example, X-linked haemophilia was shown to be linked to red green colour blindness as early as 1937. By the 1980s it became possible to locate the genes responsible for diseases inherited in a Mendelian fashion to specific chromosomes by genome mapping techniques. The development of PCR (Saiki et al. 1985) has revolutionized the field of molecular biology and the detection of genetic disorders. Correlating the inheritance of highly polymorphic markers with diseases in families has enabled the mapping of mutations to specific chromosomal regions and discovery of candidate genes (Nelson 1991).

The identification of the cystic fibrosis gene (Rommens et al. 1989, Riordan et al. 1989, Kerem et al. 1989) has demonstrated the validity of a "reverse genetics" approach (Orkin 1986) to human genetic diseases. Reverse genetics has been defined as the application of linked markers to describe the chromosomal location of a gene involved in a disease, followed by efforts to identify and isolate the gene from its map position. Using this approach a large number of disease genes have already been assigned to chromosomal regions (Gusella et al. 1983, Davies et al. 1983, Reeders et al. 1985, Bodmer et al. 1987).

To date, four of the genetic disorders in animals known to be caused by a point mutation have been localised eg: hyperkalaemic periodic paralysis in quarter horses (Rudolph et al. 1992), porcine malignant hyperthermia (Fuji et al. 1991), retinal atrophy in Irish setters (Suber et al. 1993) and DUMPS (deficiency of uridine monophosphate synthase) in cattle (Schwenger et al. 1993). In all four conditions the genes responsible were already known to cause similar disorders in other species.

"Dwarfism" in cattle is dominantly-inherited in which the homozygous condition is lethal and is particularly predominant in the Dexter breed. The mutation causing the defect has not yet been identified and efforts are being made by different laboratories to isolate the causative gene. Identification of a genetic marker associated with this defect would be of great advantage for animal breeders for early diagnosis of the carrier status. An approach was made to the identification of the mutation causing this lethal defect and is explained later in this thesis.

Different approaches are used for the identification of disease genes. 1. Cytogenetic approach, 2. Linkage analysis, 3. Genome mapping and 4. Positional cloning.

1.4.3.1 *Cytogenetic approach.*

Although in the early part of the century cytogeneticists postulated that chromosomes provide hereditary basis for certain human diseases, it was not until 1959 that a chromosomal defect was established as the cause of a specific disease in man (Basrur 1980). Since then, a number of human syndromes such as Down's syndrome, Turner's syndrome and Klinefelter's syndrome have been assigned as

specific chromosome anomalies. In domestic animals, the cytogenetic approach has attracted attention through the discovery of chromosome alterations associated with subfertility eg: 1/29 translocation in cattle (Gustavsson 1969). Chromosomal rearrangements as a cause of inherited diseases in cattle have not been as well studied as in man, but when found, they simplify the task of localizing the gene.

1.4.3.2 ***Linkage analysis.***

Linkage analysis using genetic markers has been the most powerful and practical tool in medical genetics for locating disease genes since the mid 1970s (White and Lalouel 1988, Lathrop et al. 1991). The markers used for linkage analysis must be readily detectable and be found in a number of distinguishable variants throughout the population for correlating the inheritance of a marker and disease gene. Linked-markers provide the starting point for the identification and characterisation of disease loci. Several diseases have been mapped to specific chromosomes by the method of linkage analysis in man. The examples are Huntingtons disease and cystic fibrosis (Gusella et al. 1983, Wainwright et al. 1985). In cattle, linkage studies with five DNA fingerprint probes showed a fragment associated with the muscular hypertrophy gene which is responsible for the double muscled phenotype (Georges et al. 1990a). Similarly, linkage analysis on an extensive bovine pedigree segregating for the Weaver gene identified a microsatellite locus TGLA116, closely linked to the condition which was assigned to bovine synteny group 13.

The method for analysing linkage data is the lod score method or log of odds of probability for or against linkage and is based on the work of Morton (1955) and Lathrop et al. (1984). Two loci on different chromosomes will cosegregate by chance, on average 50% of the time whereas two loci close to each other on the same chromosome will usually be inherited together on more than 50% of occasions. The frequency with which the two loci fail to segregate together, is known as the recombination fraction (θ). The likelihood method of linkage analysis is calculated by estimating the recombination fraction, and testing whether an estimate of this value is significantly smaller than 50%. To calculate the likelihood of whether the

two loci are linked or not, values of the likelihood L (θ) for different values of the recombination distance (0, 0.01, 0.05, 0.10, 0.20, 0.30, 0.40) are then compared to the likelihood of the two loci under full recombination L (0.5). The logarithm (base 10) of the ratio of the two is used and a lod score Z is known as log of odds or log probability ratio. The value of (θ) at which Z (θ) is largest is known as the maximum lod score. Linkage can be considered established when in a collection of families the sum of the lod scores at any value of (θ) reaches +3, corresponding to odds for linkage of at least 1,000:1. If, however the sum of the lod scores at a given value of (θ) is less than -2 then the linkage is considered to be excluded.

Linked-markers may be several million base pairs of DNA distant from the disease gene. In order to bracket the area of chromosome to be searched, it is necessary to identify flanking polymorphic markers. Having identified a set of markers linked to the trait of interest, the linear order relative to each other, is determined by the frequency of recombination between them caused by crossover between the marker and the disease gene. Once the most plausible order for a cluster of linked markers has been established, they are assigned to specific chromosomes.

1.4.3.3 *Genome mapping approach.*

Coverage of markers in human and mouse genomes is better than that in the bovine genome (O'Brien et al. 1993a). The chromosomal localisation of linked-markers can be achieved by various physical mapping methods. Once the chromosomal localization of the gene of interest has been determined, an inspection of the homologous regions of chromosome in man or mouse is made for any obvious candidate genes with physiological roles that are likely to be involved in the disease. If such candidate genes are found, the homologues are isolated from cDNA libraries. This candidate gene is screened for the difference in sequence between the affected and normal individuals. This is done by detecting polymorphism in the different regions of the gene using techniques like denaturing gradient gel electrophoresis (DGGE) and single stranded conformation polymorphism (SSCP) (Sheffield et al. 1989, Orita et al. 1990) which can detect even single base pair changes. On

detecting a change between the affected and unaffected group, that section of the gene is sequenced to identify the variation in the defective allele.

The gene controlling porcine stress syndrome was identified partly through mapping (Davies et al. 1988) because of linkage between the halothane locus, phosphohexose isomerase (PHI) and the H blood group locus. In cattle, the gene responsible for bovine leukocyte adhesion deficiency (BLAD) (Shuster et al. 1992) was identified by the knowledge of sequence from the corresponding locus in man.

For many disorders no obvious candidate gene has been identified. In these cases, the mutation has to be localized first to a particular region of the chromosome, then genes in the region identified by a positional cloning approach.

1.4.3.4 *Positional cloning.*

In the absence of candidate genes, the chromosomal regions identified by linkage analysis will be mapped at high resolution as a prelude to positional cloning. Initially, isolation and identification of a chromosome or region specific isolation of DNA fragments is done by somatic cell hybrid panels (Gusella et al. 1980), chromosome sorting (Gray et al. 1987) or by microdissection of specific chromosome regions (Bates et al. 1986). Positional cloning of the gene is done without any information about its protein product. The inheritance of each marker is compared with that of the disorder in families or groups of families. The markers act as guides to chromosomal locations, each marker flagging a short stretch of a chromosome. Isolation of adjacent or nearby DNA fragments in regions localised for disease gene is done by chromosome walking, by the use of a probe to identify additional cloned DNA that overlaps it, followed by preparation of probes from the farthest end of the new clone and rescreening to continue down the chromosome. Even though the technique in theory is simple, it is laborious to perform. This approach was first described by Bender et al. (1983) in Drosophila.

1.4.4 *Detection of mutation in DNA.*

Demonstration of variation is a critical step in many studies in molecular genetics. In screening for mutations in genes that contribute to a particular disease

process, direct gene analysis is the common strategy. Currently, mutations are detected by different procedures such as altered banding patterns of single-stranded DNA on non-denaturing gels (SSCP) (Orita et al. 1990), resolution of heteroduplex molecules by their instability in denaturing gradient in gels (DGGE) (Sheffield et al. 1989), cleaving DNA with chemicals (Cotton et al. 1988) or with ribonuclease A (Myers et al. 1985) or direct sequencing of the DNA segments. The recently introduced method termed enzyme mismatch cleavage (EMC) (Youil et al. 1995) is regarded as superior to the existing methods for detection of mutation. Bacteriophage resolvases, whose function *in vivo* are to cleave branched DNA, have the property of recognizing mismatched bases in double stranded DNA and cutting the DNA at the mismatch. Radiolabelled-DNA is cleaved by the resolvase at the site of mismatch in heteroduplex DNA and the digestion is monitored on a gel. Thus, both the presence and estimated position of an alteration are revealed.

The carrier status of animals for certain diseases can be detected by biochemical screening eg: enzyme alpha mannosidase for mannosidosis (Jolly and Blakemore 1973), and pyruvate kinase for congenital haemolytic anaemia in dogs (Prasse 1977). With the advances in molecular biology, DNA tests are employed to detect the carriers in the population. For instance, BLAD is a recessive disorder in cattle caused by a point mutation within the gene encoding bovine alpha subunit of CD18 and carriers are identified by the PCR-RFLP (Shuster et al. 1992). Similarly, DUMPS, a monogenic disorder in the Holstein breed (Shanks et al. 1984) has been shown to result from a point mutation generating a stop codon in UMPS. Carriers are detected by a direct DNA test based on PCR (Schwenger et al. 1993). DNA-based tests are also employed for the detection of the mutation associated with malignant hyperthermia in pigs (O'Brien et al. 1993b).

1.4.5 Aims of PhD project.

Microsatellite loci provide the unique class of markers which overcome many of the difficulties associated with other marker types (outlined in section 1.2.2.2). The use of microsatellite markers is yet to be investigated to enhance the genetic progress in cattle breeding programmes through the identification of loci affecting

traits of economic importance. Therefore, the aim of this PhD project was to evaluate their use in cattle breeding. In the first part of the thesis, the efficiency of microsatellite markers was evaluated in a) parentage determination and individual identification and b) phylogenetic analysis. Methodology for typing a large number of samples on a routine basis was also investigated. Various practical applications using these markers are described.

The second part of the thesis, was an attempt to identify a marker for the dominantly inherited monogenic lethal trait in Dexter cattle, namely, achondroplasia. A mutation in the transmembrane region of the fibroblast growth factor receptor 3 is known to cause a similar condition in man. The aim of this part of the thesis is to investigate mutations in the transmembrane region of the bovine FGFR3 gene to see if the same gene and mutations are responsible for dwarfism in Dexter cattle.

CHAPTER 2
GENERAL MATERIALS AND METHODS

This chapter describes the routine materials and techniques used in this study. The preparation of common laboratory solutions and media is given in the relevant sections.

2.1 Preparation of Genomic DNA.

2.1.1 *Blood.*

The method followed has been described by Oliver et al. (1989). Blood samples (10ml) were diluted with 1/4th volume of 0.9% saline and centrifuged for 15min at 1500rpm. The buffy coat was carefully transfered to another tube and the red cells were lysed for 15sec by the addition of 2ml of distilled water followed immediately by 0.2ml of 9% NaCl. Then 10ml of 0.9% saline was added, mixed and centrifuged for 10min at 2500rpm. The supernatant was discarded and the pellet was resuspended in 2ml lysis buffer (50mM glucose, 100mM Tris, 10mM EDTA) followed by the addition of 10ul of proteinase K (20mg/ml) and 100ul of 10% SDS. The mixture was incubated at 65^{0}C for 2hrs. Following incubation a further 2ml of lysis buffer was added and then two phenol chloroform extractions were performed for 5min. A final extraction was with 5ml chloroform for 2min. DNA was precipitated with 2.5 volumes of ethanol, pelleted by centrifugation, dried under vacuum and redissolved in 250ul of Tris EDTA (TE) buffer (10mM Tris, 1mM EDTA).

2.1.2 *Milk.*

The method for extraction of DNA from milk was similar to that described by Lipkin et al. (1993). Milk was diluted with an equal volume of 0.9% saline and centrifuged for 10min at 1000rpm. The pellet was washed three times in saline, resuspended in 2ml of prewarmed lysis buffer (10mM Tris, 100mM NaCl, 10mM EDTA, 0.5% SDS, 2% β mercaptoethanol) and incubated at 50^{0}C for 30min.

Dithiothretiol and proteinase K were added to final concentrations of 70mM and 0.2mg/ml respectively and incubated for 2hrs. DNA extraction was then carried out as described for blood. Alternatively, milk somatic cells in the pellet were lysed by successive incubations in boiling water for 5min, 50^0C for 5min respectively, and again in boiling water for 5min. These cells were resuspended in 10ml saline, centrifuged for 10min and a final suspension was made in 1ml of TE buffer.

2.1.3 ***Semen.***

The semen straws (2 X 0.5ml) were thawed, and expelled into 10ml of 15mM NaCl/10mM EDTA pH 8.0. The contents were mixed well and centrifuged at 2000rpm for 10 min, the supernatant was discarded, and the pellet resuspended in 0.5ml of prewarmed sperm lysis buffer (100mM Tris, 500mM NaCl, 10mM EDTA, 1% SDS, 2% β mercaptoethanol). Proteinase K (20mg/ml) and dithiothretiol (70mM) were added to final concentration of 0.2mg/ml and incubated for 2hrs. A further 20ul of proteinase K was added and incubation was continued for 2hrs. DNA was extracted twice for 5min with phenol/chloroform and recovered by precipitation with 2.5 volumes of ethanol (Anderson et al. 1985).

2.1.4 ***Hair.***

Most of the DNA in hair is located in the root and surrounding sheath cells. The method of Gill et al. (1985) was followed. Hair samples (50-100 numbers) were rinsed in distilled water and then treated with absolute ethanol. After drying, hairs were digested in 0.5ml of lysis buffer (10mM Tris, 10mM EDTA, 100mM NaCl) containing 50ug/ml proteinase K, 40mM DTT and 2% SDS. DNA was extracted with equal volumes of phenol/chloroform for 5min and once with chloroform for 5min, precipitated with 2.5 volumes of absolute alcohol. DNA was pelleted by centrifugation at 15,000rpm for 5min, washed in 70% ethanol repelleted, dried and dissolved in 50ul of distilled water. Particulate matter was removed by passing through a centricon column, by centrifugation for 15min. The filtrate was used for DNA analysis.

2.1.5 *Tissues.*

The method described by Sambrook et al. (1989) was followed. Freshly excised tissue (25gm) was cut into small pieces and homogenized in ice-cold PBS (140mM NaCl, 3mM KCl, 8mM Na_2HPO_4, 2mM KH_2PO_4). The fragments of connective tissue were removed by filtration. The suspended cells were washed thrice in ice-cold PBS and resuspended in 2ml ice-cold L buffer (10mM Tris, 100mM EDTA, 20mM NaCl). To this was added 2ml of L buffer followed by 20ul proteinase K (20mg/ml) and 100ul of 10% SDS and the mixture was incubated for 2-3hrs at 65^0C. Subsequent procedures of phenol/chloroform extractions and recovery of DNA were the same as for blood.

2.2 Analysis of the concentration and purity of DNA.

The concentration of DNA samples was determined by measuring the absorbance of an aliquot of 1/100 dilution in a spectrophotometer at 260 and 280nm. An optical density reading of 1 at 260nm is equivalent to 50ug DNA/ml. The ratio between the readings at 260nm and 280nm (OD_{260}/OD_{280}) provided an estimate of the purity of DNA. Pure preparations of DNA have an OD_{260}/OD_{280} value of 1.8. If contaminated with protein the value is lower. Quality of the DNA was also assessed on an agarose gel. If a single high molecular weight band was visualized following electrophoresis, the DNA extracted was assumed to be intact and not degraded, while a streak revealed that the DNA had been physically sheared. A staining in the gel slots indicated protein contamination.

2.3 Incorporation of radioactivity.

2.3.1 *Direct incorporation.*

Direct incorporation of radioactively labelled nucleotides (α^{35}S dATP and α^{32}P dCTP) was done by adding 0.002mCi of α^{35}S dATP or 0.001mCi of α^{32}P dCTP into the 10ul PCR reaction mix.

2.3.2 *End-labelling of PCR primers.*

For use in the PCR, one member of each pair of primer was polynucleotide kinase labelled at its 5'end with γ^{33}P ATP. The reaction was carried out in a 10ul volume containing 1ul of 10X PNK buffer (700mM Tris, 100mM $MgCl_2$, 50mM DTT), 1.6uM primer, 6 units of PNK, 0.008mCi of γ^{33}P ATP and 6.8ul of water and incubated at 37^{0}C for 30min. For setting up the PCR, the end-labelled primer was diluted with 30ul water and 1ul was used per reaction.

2.4 General conditions for PCR.

The polymerase chain reactions were set up containing 100-200ng of template DNA with 2ul of 10X Taq buffer (100mM Tris, 500mM KCl, 1mg/ml gelatine), 200uM each deoxyribonucleoside triphosphates (dNTP), 1.25mM $MgCl_2$ 1uM of each of the oligonucleotide primers and 0.5 units of Taq DNA polymerase enzyme (Boehringer). Total volume of the reaction was made up to 20ul with water and was overlaid with liquid paraffin to prevent evaporation. Temperature cycling was carried out in a programmable thermal cycler. Amplification was for 30 cycles with each cycle of denaturation of 1min at 94^{0}C, 1min annealing at appropriate temperature (which varied with primers) and 1min extension at 72^{0}C.

It is beneficial to optimize the PCR conditions to obtain a satisfactory product after amplification. The design of primers is important to avoid accumulation of nonspecific products and to avoid primer-dimer artifacts. Enzyme concentration is also critical: if too high, non-specific background products accumulate, and if too low, an insufficient amount of the desired product is obtained. Magnesium concentration affects primer annealing and hence, product specificity, formation of primer-dimer artifact, and enzyme activity and fidelity. Temperature and the length of time required for primer annealing and extension is important for best results.

2.5 Primer design for sequencing/PCR.

Primer concentrations between 0.1 and 0.5uM were optimal for the PCR reaction. Primers designed were usually between 18 - 25 nucleotides in length with 50 - 60% G + C composition evenly distributed. Runs of four or more of the same

base and palindromic sequences were avoided. For a primer pair the melting temperature (T_m) was estimated with a theoretical annealing temperature difference within 5^0 and between 55^0C and 65^0C. Complementarity at the 3' ends of the primer pairs was avoided to prevent self annealing. Primers were chosen as close as possible to the sequence of interest to increase the proportion in the PCR product while minimizing its size.

2.6 Electrophoresis.

2.6.1 *Agarose gel electrophoresis.*

The agarose gels were prepared in the horizontal plastic tray of the electrophoresis apparatus by sealing the open ends with autoclave tape to form the mould. The concentration of agarose used ranged from less than 1% to 3% for electrophoresing larger and smaller DNA fragments respectively.

Two grams of agarose (Sigma) were mixed with 100ml of 1X TBE (45mM Tris-borate, 1mM EDTA) and heated in a microwave until the agarose was dissolved. The solution was cooled to 60^0C, mixed with ethidium bromide (10mg/ml in water) to a final concentration of 0.5ug/ml and poured on to the mould to a thickness of 3-5mm. The combs were placed in the slot avoiding any airbubbles, and withdrawn after the gel was set and positioned for electrophoresis. For preparative gels, either ultra pure agarose or low melting temperature agarose (Nusieve GTG) at an appropriate concentration were used.

The gel was placed in a horizontal tank with one litre of 1X TBE sufficient just to cover it. An aliquot of 10ul of the DNA sample was mixed with 5ul of loading buffer (50mM Tris, 5mM EDTA, 50% glycerol, 0.1% Bromophenol Blue) and was loaded into the wells of the gel. Electrophoresis was for 1hr at 100 volts. The gel was placed on a transilluminator to view the products.

2.6.1.2 *Isolation of DNA fragments from agarose gels for preparing radio labelled probes.*

Samples run on low melting point agarose gel were used for radio labelled probes. The desired bands were excised and transfered to a preweighed 1.5ml

microfuge tube. Distilled water was added at a ratio of 3ml per gram of gel and placed in a boiling water bath for 5 min to melt the gel and denature the DNA. The boiled samples were divided into suitable aliquots (25ul).

2.6.2 *Nondenaturing polyacrylamide gel electrophoresis.*

Nondenaturing polyacrylamide gels were used for the separation and purification of double stranded DNA. The Biorad vertical gel electrophoresis system used consisted of two glass plates (19.5cm X 19.5cm and 19.5cm X 22cm) separated by 1mm spacers which were placed on the rubber gasket on the base of the apparatus and were clamped tightly on either side so that the plates were completely sealed on three sides. The gels (10%) were prepared in a proportion of 19:1 acrylamide and N'-N'-Methylenebisacrylamide by mixing 7.5ml of 40% polyacrylamide solution (38% acrylamide, 2% bisacrylamide) with 3ml of 10X TBE and 19.5ml of water. Gels were polymerised by adding 60ul of freshly prepared 25% ammonium persulphate (APS) and 25ul of N,N,N,N-tetramethylethylenediamine (TEMED), then poured quickly between the glass plates and the comb was inserted. The gel was left to set for 2hrs.

Electrophoresis was in 1X TBE buffer. DNA samples were mixed with 0.2 volumes of gel loading buffer and loaded in the gel slots. The sizes of DNA fragments were determined by comparison with known DNA sized markers including 1Kb ladder (Gibco-BRL) and plasmid pBR322 restricted with HaeIII. The gels were run at a voltage of 50-80 volts overnight.

2.6.3 *Denaturing polyacrylamide gel electrophoresis.*

Conventional 0.4mm thick 6% denaturing polyacrylamide gels containing 6M urea were used to fractionate radioactively labelled single-stranded DNA and also to separate the labelled fragments in the sequencing reaction. This vertical gel system was as described by Biggin et al (1983) and manufactured by Life Technologies, GIBCO BRL. The gels were set between two glass plates (41.5cm X 33cm, 39.5cm X 33cm) separated by 0.4mm spacers. One of the glass plates was siliconised with dimethyldichlorosilane solution (BDH) to prevent the gel from

adhering to the plate. The plates were cleaned in running water, dried and wiped with alcohol. The plates were assembled with spacers in between and sealed on three sides with sealing tape. The gel was prepared by using 60ml of 6% acrylamide gel mix, mixed with 1.5ul of 25% APS and 1.5ul of TEMED (Sigma) per ml of gel mix. The contents were mixed in a cylinder and poured between the glass plates avoiding air bubbles. The plates were clamped and the combs (sharks tooth comb) inserted. The gel was allowed to set for an hour before electrophoresis. The buffer, 1X TBE, was poured in both the upper and lower electrode tanks. The combs were removed to clean the wells and reinserted.

The sample to be loaded was mixed with 4ul of formamide loading buffer (0.025% Xylene Cyanol FF, 0.025% Bromophenol Blue, 10mM EDTA, 98% Deionised formamide) and denatured at 94^{0}C for 4min. The gels were electrophoresed at 50 watts for 3hrs. The bromophenol dye in the loading buffer was an indicator of the mobility of the DNA fragments and had a mobility similar to a 25bp fragment and the xylene cyanol dye had a mobility similar to an approximate 100bp DNA fragment. For PCR products, multiple loadings at intervals could be done such that, the products were separated and could easily be read. This enabled up to 180 samples to be analysed per gel. After electrophoresis, the glass plates were carefully separated, the gels were transfered to Whatmann 3mm filter paper, covered with Saran Wrap and dried in a gel drier at 80^{0}C. The Saran Wrap was removed, and the gels were put for autoradiography with Agfa Curix RPI film for 24-48hrs in a light tight cassette at -80^{0}C.

2.7 Automated DNA typing.

2.7.1 *Fluorescent dyes.*

The fluorescent labelled primers were included in the PCR reaction to tag the PCR product with specific dyes for analysis on the ABI sequencer. The four different dyes used for automated DNA typing, TET, HEX and JOE (green), FAM (blue), TAMRA (yellow) and ROX (red). The ROX dye is used with the internal size standard and the other dyes are used to tag the DNA fragments. The chemical name and emission wavelength for each fluorescent dye are as follows.

Dye	Chemical name	Colour	Wave length
ROX	6-carboxy-rhodamine	RED	605nm
FAM	5-carboxyfluorescein	BLUE	521nm
HEX	6-carboxy-2'4'7'4,7-hexachlorofluorescein	GREEN	556nm
JOE	2',7'-dimethoxy-4'5' dichloro-6-carboxyfluorescein	GREEN	555nm
TET	Tetra chloro fluorescein	GREEN	545nm
TAMRA	N,N,N',N'-tetramethyl-6-carboxyrhodamine	YELLOW	580nm

2.7.2 *Internal size standard.*

The internal lane (size) standards are specific DNA fragments of known sizes. GENESCAN 672 software identifies any peak labelled with ROX and computes a calibration curve to calculate the fragment sizes of unknown DNA samples in relation to the standard. The ROX labelled internal size standard used in this work was GENESCAN 2500-ROX, a Pst1 digest of lambda phage. For this standard the fragments resolved are of sizes 37, 94, 109, 116, 172, 186, 222, 233, 238, 269, 286, 361, 470, 490 and 536bp.

2.7.3 ***Preparation of gels for electrophoresis in ABI.***

In the automated gel electrophoresis system, the laser scans through the gel, therefore the glass plates and the gel must be non fluorescent. All the solutions used were filtered to remove any particulate matter and the glass plates were kept absolutely clean and free of dust. The glass plates (42cm X 25cm) and the spacers (0.4mm) were aligned and sealed on three sides with the gel sealing tape. The 6% acrylamide urea gel solution was made freshly from 30gm of urea dissolved in 22ml of water and 9ml of 40% acrylamide solution. The solution was filtered through a

0.2um cellulose acetate filter into a flask containing 6ml of filtered 10X TBE buffer and was degassed for 5min. The final volume was adjusted to 60ml to which 300ul of 10% APS and 27ul of TEMED was added, mixed and the gel poured. The gel was then set for a minimum of 2hrs before the samples were loaded.

2.8 Cloning PCR products.

The TA cloning system from Invitrogen provides a cloning strategy for direct insertion of a PCR product into a plasmid vector. The TA cloning system takes advantage of the non-template dependent activity of thermostable polymerase used for PCR that adds single deoxyadenosines to the 3' end of all duplex molecules. These A-overhangs are used to insert the PCR product into a specifically designed vector (pCR™) providing 3' T-overhangs at the insertion site.

The method of cloning the PCR product was as described in the TA cloning instruction manual version 1.3 by Invitrogen corporation, San Diego, California. The ligations with the pCR™ vector (3.9Kb) were set up as 1:1 to 1:3 molar ratio of the vector:PCR insert.

2.8.1 ***Ligation reaction.***

Ligation reactions were set up with 6ul sterile water, 1ul of 10X ligation buffer (200mM Tris, 50mM $MgCl_2$, 50mM DTT, 500ug/ml BSA, 1mM ATP,) 2ul pCR™ vector (25ng/ul), 1ul of PCR product and 1ul of T4 DNA ligase. The mixture was incubated at 13^0C overnight.

2.8.2 ***Transformation of ligations.***

The ligation reaction was mixed briefly and centrifuged and placed on ice. β mercaptoethanol (BME) 0.5M was added to a 50ul vial of thawed competent *E.coli* cells (New England Biolabs) and mixed by tapping. The ligation reaction (1ul) was added to the competent cells and incubated on ice for 30min, heat shocked at 42^0C for 1min then chilled on ice for 2min, followed by addition of 450ul of prewarmed SOC medium (2% Bactotryptone, 0.5% Bacto yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM Glucose). The cells were

incubated at 37^{0}C for 1hr in a shaking incubator to allow for expression of antibiotic resistance gene and then plated on to fresh ampicillin (50ug/ml) LB agar plates spread with 25ul of X Gal (40mg/ml). Transformed colonies were observed after incubation at 37^{0}C overnight.

2.8.3 *Identification of recombinant plasmids.*

Identification of the recombinant plasmids was performed by the alpha complementation test. Most of the colonies carry a short segment of *E.coli* DNA that contains the regulatory sequences and coding information of the first 146 amino acids of the beta galactosidase gene (lac Z). A polycloning site within this region insert aminoacids into amino terminal fragment of beta galactosidase. Vectors of this type are used in host cells that code for the carboxy terminal portion of beta galactosidase. Host and plasmid encoded fragments associate to produce an enzymatically active protein. In alpha complementation, proximal segment of lacZ gene are complemented by betagalactosidase. The lac$^+$ bacteria that result from alpha complementation are recognised by blue colonies in presence of 5-bromo 4-chloro-3 indolyl- β O-galactosidase (X-Gal). Thus bacteria that form blue colonies in presence of X Gal are with non-recombinant plasmids and those forming white colonies are with recombinant plasmids. The structure of these plasmids were verified by restriction analysis of minipreparations of plasmid DNA or by PCR colony miniprep.

2.8.4 Preparation of plasmid DNA

2.8.4.1 *Minipreparation of DNA.*

An overnight culture of the bacteria containing plasmid was grown in 5ml of LB medium with ampicillin (100ug/ml). The cells were harvested by centrifugation at 3,000rpm for 10min and the supernatant was discarded. The purification of plasmid DNA was performed using the Wizard miniprep DNA purification system (Promega). The cells were resuspended in 200ul of cell resuspension solution (50mM Tris, 10mM EDTA, 100ug/ml RNase) to digest RNA. The cells were then lysed in 200ul of cell lysis solution (0.2M NaOH, 1% SDS) and then neutralised in

200ul of neutralisation solution (1.32M Potassium acetate - pH4.8). The mixture was centrifuged and the supernatant was mixed with 1ml of DNA purification resin. The slurry was passed through the minicolumn attached to a syringe. The column was washed twice with column wash solution (200mM NaCl, 20mM Tris, 5mM EDTA: diluted 1:1 with 95% ethanol) and added 50ul of water. The minicolumn was centrifuged in a microfuge to elute the DNA.

2.8.4.2 ***Maxipreparation of DNA.***

Large scale preparation of DNA from an overnight culture of bacteria in 500ml LB medium was performed using Qiagen plasmid kit (Qiagen). The culture was centrifuged at 8,000rpm for 10min and the bacterial pellet was resuspended in buffer P1 (resuspension solution) to digest the RNA. To this, 10ml of buffer P2 was added (cell lysis solution) and the mixture was incubated at room temperature for 5min to lyse the cells and 10ml of buffer P3 (cell neutralisation solution) was added and the mixture was incubated on ice for 20min to ensure the precipitation of protein and high molecular weight DNA. The mixture was then centrifuged for 30min at 4^0C and the supernatant was passed through a Qiagen tip 500 column after equilibriating with buffer QBT (750mM NaCl, 50mM MOPS, 15% ethanol, 0.15% Triton X 100). The column was washed with buffer QC (1M NaCl, 50mM MOPS, 15% ethanol) and the DNA was eluted with 15ml of buffer QF (1.25M NaCl, 50mM MOPS, 15% ethanol). DNA was precipitated with 0.7 volume of isopropanol, collected by centrifugation for 30min, washed with 10ml of 70% ethanol and redissolved in 250ul of TE buffer.

2.9 Restriction endonuclease digestion of DNA.

Restriction endonuclease digests of plasmid or phage DNA were carried out in 20ul reaction volume following the manufacturer's instruction. Typically 5-10ug of genomic DNA or 0.5ug of plasmid DNA were digested with 5-20 units of enzyme for 1-2hrs at 37^0C. Small aliquots of the digest were checked on agarose mini gels for complete digestion (section 2.6.1). For better resolution of small fragments after digestion, the samples were electrophoresed on nondenaturing polyacrylamide gels.

2.10 Polymerase Chain Reaction (PCR) colony miniprep.

Recombinant vectors were analysed directly by PCR from the bacterial cells. The protocol followed was as described by Hawker and Billadella (1993). A bacterial colony was touched with a sterile pipette tip or 1ul of an overnight culture was inoculated into 100ul of sterile distilled water. The sample was boiled for 1min and chilled on ice for 5min. Then 80ul were withdrawn and added to 10ul of 10X PCR buffer, and 10ul of 2mM stocks of dNTPs, 1uM of each PCR primer flanking the insert site in the vector and 2.5 units of *Taq* DNA polymerase (GIBCO) in a total volume of 100ul. The sample was amplified by PCR with an initial cycle of 94^{0}C for 5min, 55^{0}C for 4min and 72^{0}C for 4min followed by 25 cycles of 94^{0}C for 1min, 55^{0}C for 2min and 72^{0}C for 4min. After PCR 15ul of each reaction were analysed by electrophoresis through a 2% agarose gel to identify the colony with the insert of the correct size.

2.11 Purification and concentration of DNA.

If the prepared DNA was too dilute, precipitation with glycogen and ethanol was performed as described by Sambrook et al. (1989). DNA (100ul) was mixed with 1/10th volume of glycogen (10mg/ml) and 3M sodium acetate pH 4.7 and 2.5 volumes of 100% ethanol, incubated at -20^{0}C for 30min and then centrifuged at for 20min in eppendorf microfuge. The ethanol supernatant was removed, air dried for 5-10min and dissolved in 10ul of water.

2.12 Sequencing of plasmid DNA.

DNA sequencing was performed by the chain termination method of Sanger et al (1977). This method employs 2'3' dideoxynucleoside triphosphates (ddNTP) to terminate growing DNA strands. Denatured double stranded DNA was annealed to an oligonucleotide primer, which was homologous to the 5' side of the sequence of interest. Sequenase enzyme was used to extend from the primer by its 5'-3' polymerase activity. *In vitro* synthesis of DNA was carried out in four separate incubation mixes in the presence of one of the four ddNTPs which acts as a chain terminator whenever incorporated. This is because, on dideoxynucleotide, the 3'

hydroxyl group necessary to form the next phosphodiester bond during DNA synthesis is absent. Thus in each reaction, nested sets of molecules were generated, a proportion of which ended each time there was a nucleotide in the sequence corresponding to a particular ddNTP (Sanger and Coulson 1975). The fragments were then denatured and resolved in thin (4mm) 6% denaturing polyacrylamide gels (section 2.6.3).

2.12.1 *Denaturation of plasmid DNA.*

Denaturation of plasmid DNA was done as described by Zhang et al (1988). Approximately 10ug of double-stranded circular plasmid DNA was denatured by adding 2ul of 2M NaOH/2mM EDTA to 20ul of DNA and incubated at room temperature for 5min. The mixture was neutralised by adding 2ul of 3M sodium acetate (pH 4.5-5.5) and the DNA precipitated by adding 2-3 volumes of 100% ethnol and incubated at -20^{0}C for 20min. The precipitated DNA was collected by centrifuging at 12,000rpm for 20min and the pellet washed in 70% ice cold ethanol, dried and resuspended in 7ul of distilled water to be used in annealing reaction.

2.12.2 *Sequencing using Sequenase kit.*

Sequenase^R version 2.0 sequencing kit (United States Biochemical, Cambridge, UK) was used to sequence the denatured DNA as per manufacturer's instructions. The reaction was carried out in three steps: annealing, elongation and termination. In brief, to the denatured DNA (7ul) 2ul of 5X reaction buffer (200mM Tris, 100mM $MgCl_2$, 250mM NaCl) and 1ul of the primer of interest (1uM stock) were added and heated for 2min at 65^{0}C and cooled slowly to less than 35^{0}C. After annealing, 1ul of 0.1M DTT, 2ul of 1:9 dilution of labelling mix (7.5uM dGTP, 7.5uM dCTP, 7.5uM dTTP), 0.5ul of α^{35}SdATP (Amersham) and 2ul of 1:7 dilution of sequenase version 2 enzyme (diluted in enzyme dilution buffer - 10mM Tris, 5mM DTT, 0.5mg/ml BSA) were added to the tube and incubated at room temperature for 3-5min. 3.5ul of this labelled reaction were dispensed into each of the four tubes (labelled A, C, G and T) containing 2.5ul of one of the four termination mixes (ddA, ddC, ddG and ddT) and incubated at 37^{0}C for 5min.

Reactions were stopped by the addition of 4ul of stop buffer.

2.12.3 *Cycle sequencing.*

The thermal cycle sequencing reaction protocols are based upon the dideoxy nucleotide chain termination method of Sanger et al (1977), but repetitive cycles of denaturation, annealing and chain extension allow only small amounts of template to be sequenced. Vent DNA polymerase (New England Biolabs), dNTPs and ddNTPs achieve an amplification of the reaction products and strong sequencing signals. The thermal cycle sequencing differs from sequenase reaction by 1. using less template DNA for the reaction, 2. using double-stranded DNA as template and 3. absence of a separate annealing reaction.

The procedure followed was as described in the cycle sequencing kit by New England Biolabs Ltd. In brief, 200ng of double-stranded DNA were mixed with template, 3ul of 1uM primer, 1.5ul of 10X circumvent^{TM} sequencing buffer (200mM Tris, 100mM KCl, 100mM $(NH_4)_2SO_4$, 50mM $MgSO_4$), 1ul of 30X Triton X-100 (3% Triton X-100), 0.02mCi of α ^{35}SdATP, 1ul of Vent$^{TM}{}_R$ (exo$^-$) DNA polymerase (2 units) and water to make up to 15ul. 3.2ul of the reaction mix was added to the microfuge tubes containing 3ul of ddATP, ddCTP, ddGTP and ddTTP mix, and the tube overlaid with a drop of liquid paraffin oil. The tubes were placed in a thermal cycler and the PCR amplification protocol was a denaturation at 95^{0}C for 5min followed by 20 cycles of 95^{0}C for 1min, 60^{0}C for 1min and 72^{0}C for 1min. The reaction was stopped by adding 5ul of stop buffer. The reaction was stored at +4^{0}C until analysed on 6% denaturing polyacrylamide gels.

2.13 Screening of Lambda gt11 library.

2.13.1 ***Bacterial strains*** 1.*E.coli* strain INV α F'*(endA1, recA1, hsoR17(r^k,m^{+k}), supE44*, λ-, thi-1, gyrA, relA1, Φ80lacZ$\alpha\Delta$M15Δ(lacZYA-argF), deoR+,F') was used for subcloning. The 80d lacZ M15 marker provides complementation of the beta galactosidase gene from PUC or similar vectors and, therefore, can be used to produce blue/white screening colonies on bacterial plates containing X gal as described before in section 2.8.3.

2. The E.coli strain Y1088 (Δ(lac) U169, supE, supF, hsdR, mcrA, metB, trpR, tonA21, proC::Tn5 (kanr)) was used for propagating lambda gt11 bacteriophage (λlac5 Δshn dIIIλ2-3 srIλ3^0 cIts857 srIλ4^0 min5 srIλ5^0 Sam100).

2.13.2 ***Preparation of plating bacteria.***

The bacterial strain, Y1088 was streaked on an LB agar plate and incubated overnight and a single bacterial colony was used to inoculate 50ml of NZY broth supplemented with 0.2% maltose. The culture was grown at 37^{0}C until mid log phase (OD_{600} of 0.5). The cells were then harvested by centrifugation at 3,000rpm for 10min at room temprature, the supernatant was discarded and the cell pellet was resuspended in 20ml of sterile 0.01M $MgSO_4$. The plating cells were stored at 4^{0}C for upto 3 weeks.

2.13.3 ***Plaque assay.***

Serial dilutions of the bacteriophage vector were made in SM buffer (50mM Tris, 8mM $MgSO_4$, 100mM NaCl, 0.01% Gelatin). The bacteriophage diluent (100ul) were mixed with 0.2ml of plating bacteria, incubated at 37^{0}C for 20min, added to 2.5ml top agarose (0.7% Agarose in L broth) (47^{0}C). The contents were poured onto 90mm NZY agar plates and incubated overnight for the plaques to develop. If the titre of the bacteriophage was too high, complete lysis occured and if too low, very few plaques were seen. In plates where plaques were well separated, the number of plaques were counted and the titre was determined as pfu/ml.

2.13.4 ***Transfer of bacteriophage plaques on nitrocellulose filters.***

Aliquots of bacteriophage in SM buffer containing no more than 50,000 plaques in a volume of 100ul (determined by plaque assay) were mixed with 0.3ml of plating bacteria (section 2.13.2) and incubated at 37^{0}C for 20min. Then 2.5ml of molten (47^{0}C) top agarose was added to the mixture and poured on a dry 90mm agar plate, then incubated at 37^{0}C overnight until the plaques were 1.5mm in diameter. The plates were chilled at 4^{0}C to harden the top agarose and nitrocellulose

filters (Amersham) were numbered and placed on the surface of the top agarose for 1min (without trapping air bubbles) to allow the transfer of bacteriophage. Filters were marked at 3-4 locations by stabbing through it to the agar with a needle dipped in black drawing ink. The filters were transfered to a filter pad soaked with denaturing solution (0.5M NaOH, 1.5M NaCl) left for 5min and then transfered to neutralising solution (0.5M Tris, 1.5N NaCl), for 5min. The filters were rinsed in 2X SSC (0.3M NaCl, 0.03M Sodium citrate) and dried on paper towels. The DNA was fixed by UV crosslinking. Hybridisation of filters was done as described in section 2.14.2.

2.14 Southern transfer of DNA fragments from agarose gels to Hybond-N filters.

A modification of Southern's method was used for transferring DNA fragments into Hybond-N filters (Southern 1975). To aid transfer of larger DNA fragments (greater than 10kb), the DNA in the gel was placed in 0.25 M HCl for 10min. This procedure, known as depurination, breaks up large DNA fragments in the gel by cleaving the DNA helix at random purine bases throughout its length, the resulting smaller DNA fragments to be more easily transferred onto the nylon membrane. The gel was then placed in denaturating solution for 40min with constant shaking to denature the DNA. The gel was then immersed in neutralizing solution for 40min.

A glass plate large enough to hold the gel was placed over a trough that contained 20X SSC. A wick of Whatman 3mm blotting paper was placed over the plate with both ends dripping into the 20X SSC. The gel was placed on the blotting paper and covered with a piece of "Hybond-N" filter (Amersham) prewetted in 2X SSC. Care was taken not to trap air bubbles and a 10ml graduated pipette was used as a rolling pin to squeeze out any trapped bubbles.

Four pieces of Whatman 3mm paper the size of the gel were placed on top of the Hybond-N filter. A large stack of paper towels (approximately 10cm high) was then placed on top of the 3mm paper sheets. The stack was in turn covered by a sheet of glass and 1kg weight. Polythene ("cling film") was placed around the gel

to prevent any "short circuiting" of the wicks straight through to the paper towels. Transfer was allowed to take place for 16-24hrs. The Hybond-N filter was then blotted dry and baked in a oven for 2hrs at 80^0C and placed on a UV transilluminator for 5min, to crosslink the DNA to the membrane.

2.14.1 *Preparation of ^{32}P labelled probes.*

^{32}P dCTP probes were prepared using the megaprime DNA labelling kit (Amersham). Practically all sequence combinations are represented in the primer mixture, which leads to more or less random copying of a single-stranded DNA template by Klenow enzyme (DNA polymerase 1 protein fragment having polymerase activity but no exonuclease activity). This results in an equal degree of labelling along the entire length of the probe.

To an aliquot (25ul) of the melted agarose gel slice with the DNA fragment of interest (section 2.6.1) was added to 5ul of primer and denatured for 5min at 99^0C. 10ul of megaprime reaction buffer (dATP, dCTP, dGTP, dTTP, Tris, $MgCl_2$, β mercaptoethanol), 2ul of Klenow enzyme and the denatured mixture were added to 0.1mCi of α ^{32}P dCTP (in pink dye) and incubated at 37^0C for 30min. The unincorporated radioactive nucleotides were removed from the labelled DNA by passing through a Sephadex G-50 plastic column (Pharmacia).

The labelling mix was loaded on to the Sephadex G-50 plastic column equilibriated with TE buffer, and was washed twice with 400ul of TE buffer and the elute was collected in separate eppendorf microfuge tubes. DNA containing radioactive nucleotides migrates well ahead of the unincorporated radionucleotides and could be easily recognised, using a Geiger counter, as the first peak of high radioactivity. The second peak represented the unbound nucleotides migrating near the pink dye. There is a sufficient gap between the two peaks to collect the highly radioactive DNA in eppendorfs without contamination of the unbound nucleotide fraction.

Specific activity of the labelled DNA was calculated using the formula

[% of incorporation/total amount of DNA(ng)] X 10^3 dpm/ug

2.14.2 *Hybridization of Southern blots to radioactive probes.*

Hybridisation was carried out in glass cylinders in a "Hybaid" hybridisation oven. The Hybond-N filter from the blotting process was placed in a hybridisation cylinder and prehybridized, at 65^{0}C, in a shaking incubator, with at least 15ml of Church and Gilbert hybridization solution (6.7% SDS, 0.5M Na_2HPO_4 and 10% dextran sulphate). After prehybridization the buffer was removed from the cylinder. The labelled double-stranded DNA probe was denatured by boiling in a waterbath for 5min before being placed immediately on ice to stop excessive reannealing of homologous strands. The denatured DNA probe was then added to 15ml hybridization solution which in turn was immediately mixed and added to the cylinder. The lid was closed tightly and was incubated overnight for 16-24hrs at 65^{0}C in a rotary incubator.

2.14.3 *Washing of Southern blots before autoradiography:*

After hybridization was complete, the filter was removed from the glass cylinder and washed twice for 10min at room temperature in 250ml of 2X SSC with gentle shaking. The filter was then placed in a preheated solution of 0.5X SSC/0.1% SDS and washed at 68^{0}C twice for 15min. Further 0.1X SSC/0.1% SDS washes were sometimes required depending on the background and stringency level (i.e. nearly zero), compared to actual lane signal.

2.14.4 *Autoradiography of Southern blots.*

The filter was inserted into a heat sealed bag and placed into a light sealed cassette with or without an intensifying screen depending on the intensity and/or resolution of the signal required. An X-ray film was placed on top of the filter under dark room conditions and exposed at -80^{0}C overnight before developing.

CHAPTER 3

EVALUATION OF MICROSATELLITE MARKERS FOR PARENTAGE VERIFICATION IN CATTLE

3.1 Introduction

3.1.1 Parentage determination in cattle.

Improvement of domestic livestock for traits of economic importance is dependent upon selection of future breeding stock whose performance is better than that of the population average. The genetic merit of a sire is largely determined by evaluating his progeny. The value of such progeny test information depends on the accuracy of the parentage records (Van Vleck 1970).

As a result of the rapid development of artificial insemination in cattle, the practice of mating cows to different bulls during the same or two consecutive heat periods has increased the risk of unknown or uncertain pedigree. When the oestrus cycle is shorter than normal or resulting calving is earlier than the predicted dates, it may not be possible to decide the actual sire of the new born calf on the basis of records. Exchange of semen samples and misidentification of semen used are also possible stages where errors occur. Poor record keeping and wrong entry of dam identity of the newly born calf may also lead to disputed parentage, particularly in farms where weaning at birth is practised. Cases of uncertain parentage also occur often in herds practising natural service particularly when several bulls are present. Dishonest mistake could also be made when the breeder wants the pedigree to look better than it actually is. The risk of such errors makes it necessary to establish the parentage of individual animals to verify the pedigrees of each animal. Also the breed societies/herd books require verification of parentage on a routine basis.

By the beginning of 20th century, when Mendel's laws were found applicable to human blood groups, animal geneticists studied the importance of blood group genetics in livestock species (Rendel 1958). The application of blood group studies for parentage verification in cattle has been discussed extensively (Rendel 1958, Stormont 1958, Hall 1959 and Spooner 1967). Although coat colour was used for

parentage verification in cattle in the early days, blood groups were the first set of markers widely accepted for parentage tests. In addition to the blood groups, advances in electrophoretic separation methods and immunochemistry enabled analysis of soluble proteins giving additional genetic markers which could be used to increase precision. Thus for the past 30 years parentage verification has been based on readily identified gene products. These are discussed in detail later (section 3.1.2 & 3.1.3). Recently, DNA polymorphisms, have come to the forefront of interest for parentage verification, initially in man (Jeffreys et al. 1985) and more recently in livestock species (Georges et al. 1994). Validation and evaluation of DNA markers for parentage verification is the first topic addressed in this thesis.

3.1.2 Blood group serology.

The need to confirm family relationships and to resolve disputed parentage can be traced back to biblical days. The first published case of disputed maternity was the one resolved by Solomon (reviewed in Silver 1989). The Chinese employed a unique blood test to solve paternity problems in the 12th - 13th century. One method required the dripping of blood from a relative onto the skeleton of the deceased. If the blood was absorbed into the bone the relationship was considered correct. When two living individuals were involved, blood from each was dripped into a bucket of water. If the drop of blood flowed together and mixed a relationship was accepted (Silver 1989). The tests employed now are more sophisticated and employ precise immunological methods.

Following the discovery of the ABO blood groups in man in 1900 by Landsteiner, Ottenberg in 1921 proposed the use of blood grouping in genetic analysis of parentage (reviewed by Wenk et al. 1992). The description of the genetic basis for the inheritance of the ABO blood groups by Bernstein in 1924 paved the way for the use of red cell antigens as genetic markers for paternity testing in man (reviewed by Silver 1989). In farm animals, parentage tests with the aid of blood grouping were first used for horses by Kaempffer 1935 (reviewed by Rendel 1958). Blood group markers have been widely used for paternity testing in cattle for the past 20-30 years (reviewed by Stormont 1978).

The fundamental principle of all tests for disputed parentage with the aid of a genetic marker is that the factor carried by an individual must be present in either or both of its parents. When it is present the stated parentage *may be* correct. When a factor is present in an individual but not in either of its parents the given parentage must be wrong. Blood grouping has been established as an indisputable tool for parentage confirmation in cattle (Stormont 1967) and is thought to give up to 97% accuracy ie. 97 of 100 cases can be resolved (Spooner pers comm).

3.1.2.1 Red cell antigens.

Cattle red cells consists of a bilipid plasma membrane associated with protein and carbohydrate. There are slight differences in the structure of these proteins or their associated carbohydrate between individual cows which form the red cell antigens. These protein variants are detected by immunological tests. Blood from one individual is injected into another so that the recipient produces antibodies reacting with red cell antigens of the donor which are different from its own. After separation into different types/specificities these antibodies can then be used to detect the antigens on the erythrocytes of other individuals and are termed blood typing reagents. Cattle erythrocytes do not agglutinate easily. Therefore the reactions are detected in a haemolytic test (Hall 1959) in which a suspension of red cells is mixed with antibody and complement (fresh serum from rabbit or guinea pig). Binding of the antibody to the red cell antigen in the presence of complement results in lysis of the cell and the release of haemoglobin (Spooner 1980).

Serum from immunized animals may contain haemolytic antibodies directed against a few or more red cell antigens. Such a serum is known as polyvalent antiserum. Serum having antibody against a specific red cell antigen or monovalent antiserum is obtained by absorption of polyvalent antiserum with cells which have the antigen recognised by antibody. The resulting antiserum is tested using haemolytic test on known antigens and then used as a blood typing reagent. Thirty to 40 monovalent sera are used in cattle for undertaking parentage confirmation and identification of individual animals.

3.1.2.2 Inheritance of blood groups.

The antigenic moiety/determinant detected by a monospecific reagent is termed a blood group factor or a specificity. Each specificity is encoded by a single gene. A group of factors coded by linked genes occur as a unit and is termed as a blood group system, those on an individual chromosome are refered to as phenogroup. The blood group phenotype is a combination of phenogroups inherited from both parents. Phenogroups can vary occasionally through recombination. The B and C blood group systems in cattle are coded by closely linked genes. The chromosomal location of different blood group systems have been determined (Table 3.1) (Womack et al. 1992).

3.1.2.3 Blood group systems in cattle.

Rendel (1958) estimated that in Swedish Friesians only two individuals in 10,000 would be expected to have identical combination of types of blood antigens because of the large number of combinations possible. Eleven blood group systems have been recognised in cattle (Table 3.1) (Neimann-Sorrensen 1956). Some of the systems, namely the J and L are simple with only one antigen presented at each, others are more complicated, especially the B and C system (Stormont 1967). The J system in cattle is the only one to be associated with a naturally occuring antibody - anti J (Stone 1962) and the titre of anti J varies greatly according to season of the year and the age of the animal (Hall 1959). The FV locus was thought to have only two alleles - F and V, however later work suggested that there are more specificities occuring in the system (Ross and Larsen 1981, Hall and Ross 1981, Larsen 1982). The antigen of the M system is known in two forms - M_1 & M_2 one of which is dependent on the presence of the other, and is thus called a subtype (Stormont 1978). The A, S and C systems have more antigens than J, L and F systems with several subtypes.

The B system is the most extensive blood group system, for which about 300 variants have been reported. The fascination of this system is that all phenogroups are not equally present in different breeds or herds of cattle. Some of the blood groups are characteristic of certain breeds, while others are common to many. For

Table 3.1 Blood groups in cattle

Locus	Chromosome	Linkage group	Synteny group	Antigens	Reference
A	15	5	19	ADHZ'	Womack et al. (1992)
B		12		$BGI_1I_2KO_1O_2O_3O_4P_1P_2Q_1Q_2$ $T_1T_2Y_1Y_2A'B'D'E'_1E'_2E'_3$ F'G'I'$J'_1J'_2$K'O'P'Q'Y' B"G"I"	Georges et al. (1990b) Grosclaude et al.(1990) Hines&Larsen(1990)
C				$C_1C_2ER_1R_2WX_1C'L'$	
F/V		7		F_1V_1	"
J		2	16	J	Hines et al. (1969) Womack et al. (1992)
L				L	
M	23	1	20	MM'	Leveziel&Hines(1984) Lindberg&Anderson(1988)
S		7		$S_1U_1U_2H'U'_1U'_2H"S"U"$	Georges et al. (1987) Georges et al. (1990b) Grosclaude et al.(1990) Hines&Larsen (1990)
Z	8	12	18	Z	Georges et al. (1990b) Grosclaude et al.(1990) Hines&Larsen (1990) Larsen et al. (1992)
R'/S'				R'_1S'	
T'				T'	

instance, the blood group R_1 is common in Herefords and rare in other breeds (Stormont 1967) while the B phenogroup OTE'_3K' is common in Jersey and unknown in Friesians (Spooner 1967). Thus some phenogroups are indicative of breeds. Each animal inherits one phenogroup from each parent, so the blood type as it appears is a combination of two contributions. An animal which has the blood type $GY_1D'O_3J'K'$ can be formed by two B phenogroups, which are distinguished by segregation in calves, half of which may inherit the blood type GY_1D' while the other half inherit $O_3J'K'$. The B system provides a large number of different genetic markers which greatly increases the sensitivity of blood group methods in discriminating between different population of cattle.

3.1.3 Protein polymorphism.

Apart from antigenic variation between the red cells of different individuals there are variations among serum proteins. These variants are normal with respect to their biological function and differ from each other in a few amino acids. The variation in the serum proteins can be detected by electrophoresis. The electrical charge of the protein molecule depends on its amino acid content. Amino acid substitutions occurring in proteins as a result of mutation alter the charge of the protein, which affects their electrophoretic migration.

The first electrophoretic protein variant described was S haemoglobin (Pauling et al. 1949). Haemoglobin and transferrin are frequently used for parentage verification in domestic animals. Transferrin is a protein which transports iron in the body and integrates this on to the haemoglobin in the reticulocyte (Richardson et al. 1973). Allelic differences are revealed in the polypeptide chains of transferrin molecule depending on the amount of sialic acid attached (Stratil and Spooner 1971, Richardson et al. 1973). Four alleles namely A, D_1, D_2 and E are described for the transferrin protein of British cattle (Spooner 1967). More alleles occur in other breeds such as G and F in zebu cattle (Jamieson 1965). Most commonly studied proteins other than transferrin and haemoglobin are albumin, amylase, ceruloplasmin, esterases and alkaline phosphatase (Spooner 1980). The efficiency of parentage testing based upon blood groups is found to increase by 5% when

supplemented with information on transferrin and haemoglobin polymorphisms (Stormont 1967).

3.1.4 Disadvantages of blood typing and protein polymorphisms.

The accuracy of conventional blood typing for parentage verification is limited due to the finite number of blood group systems (11) and the level of polymorphism of these systems. In addition, developing a complete set of reagents for the red cell antigens and the standardisation of the technique is difficult and time consuming. Although a large number of serum proteins are available, many of them are present at low concentration and are difficult to detect (Mc Dermid et al. 1975) or have limited polymorphism (Spooner 1980), while the presence and quantity of protein at a few loci are affected by age and disease. The amount of total haemoglobin is reduced in diseases such as babesiosis thus affecting concentration of erythrocytes. Moreover, the presence of null alleles often complicates interpretation of the results as they are recessive to all other alleles and the phenotype does not correspond to the genotype (Nicholas 1987).

3.1.5 DNA based technology.

To overcome the technical limitations in serological methods and protein polymorphisms, DNA based technologies have been introduced in forensic medicine (Jeffreys et al. 1991a, Wenk et al. 1992). The large number of DNA polymorphisms present in the genome allow a large number of loci to be used, which increases the precision of parentage determination. DNA typing methods theoretically achieve a better than 99.9% probability of excluding incorrect parentage, and offer a high level of precision in individual identification (Glowatzki Mullis and Fries 1994, Vankan et al. 1994, Usha et al. 1995).

DNA based technologies are very useful in parentage determination in a herd where blood samples cannot be used. For example, in diseases such as foot and mouth disease (FMD) sources of DNA other than blood can be used eg: semen, meat or carcass. Putrified blood samples that cannot be used for routine blood typing can also form the source of DNA for parentage determination. Genotyping

of dead individuals is possible because DNA can be extracted from all nucleated cells. Semen straws stored for a long time enable the retrospective genotyping of sires. DNA based systems are not affected by age of the animals. Moreover, DNA can be stored for an unlimited period of time, thus allowing the DNA banks to be set up.

Restriction Fragment Length Polymorphisms, and the Variable Number of Tandem Repeats loci are better markers than blood groups and protein polymorphisms for parentage verification as explained in chapter I. Low levels of polymorphism and the technical difficulties involved in typing of RFLPs (Allen et al. 1990) make such markers unsuitable for routine parentage testing (Wenk et al. 1992). Even though minisatellites are more polymorphic markers than RFLPs, the hypervariable nature of the loci together with laboratory techniques involved in the analysis limits their use in parentage verification. In contrast, microsatellites represent a codominant system allowing consideration of both alleles and therefore provide complete genotype information for each locus. Microsatellite markers are ideal for parentage testing because of their high level of polymorphism and relative ease in the analysis of samples.

In recent years there has been considerable effort made to isolate highly polymorphic markers in cattle, horses, pigs, sheep and chicken. There are now more than 700 well characterised microsatellite loci in cattle, about 300 in pigs and less than 100 in sheep and chickens (Barendse et al. 1994). Sequence conservation across closely related species allows markers to be shared between cattle, sheep, goat and buffaloes (Moore et al. 1991). Microsatellites in the bovine genome are discussed in detail by Barendse et al. (1994), Bishop et al. (1994), Vaiman et al. (1994), Moore et al. (1994), Georges et al. (1995).

Microsatellite markers have been suggested as an alternative to blood typing for parentage determination in cattle, goat, pigs and dogs (Glowatzki-Mullis et al. 1995, Amigues et al. 1994, Hohenhorst et al. 1994, Mellersh et al. 1994, Usha et al. 1995). DNA typing in cattle was considered for the first time in 1993 in the International Comparison Test where the panel of markers and techniques involved in bovine paternity tests were compared between laboratories but with poor results

(Caldwell pers comm). Considerable progress has been achieved in the automation of bovine paternity testing with microsatellites using fluorescent technology (Ziegle et al. 1992, Vankan et al. 1994, Glowatzki-Mullis et al. 1995). The application of automated DNA typing in bovine paternity testing is explained in chapter 4.

In the present study five markers were selected from the available panel of microsatellite markers on the basis of their polymorphism and ease of typing. The gene coding for the bovine major histocompability complex class II antigen, DRB3, has been sequenced and a microsatellite region revealed within the 2nd intron (Muggli-Cockett and Stone 1988). The DRB3 microsatellite is composed of three repeat motifs, a stretch of at least 10 uninterrupted $(TG)_n$ dinucleotides, a long but interrupted stretch of $(GA)_n$ dinucleotides, and a few $(CAGA)_n$ tetranucleotides (Ellegren et al. 1993). A second linked gene, steroid 21-hydroxylase, CYP21 also contains a microsatellite sequence, within the 6th intron (Chung et al. 1986). The repeat sequence is a stretch of $(CA)_{20}$ dinucleotides. In cattle, both DRB3 and CYP21 loci are present on chromosome 23 (Fries et al. 1986), and are separated by only 1.4 cM on the latest map of the bovine genome (Barendse et al. 1994). Both markers have been suggested for use in pedigree verification (Penedo and Medrano 1992).

Up until now the usefulness of the markers has been considered independently and it has not yet been shown whether these tightly linked markers are in linkage disequilibrium at the population level. Unless this is established it cannot be determined whether the use of such linked markers would reduce information through redundancy, or whether gametic association would add greater power to the analysis in comparison with similar unlinked markers.

A microsatellite sequence is also found within the beta subunit of follicle stimulating hormone -FSHB which has a stretch of $(AT)_{20}$ repeats (Kemp and Teale 1991) and is found to be highly polymorphic. Microsatellites within the DRB3, CYP21 and FSHB genes together with the anonymous sequences D21S4 - ETH131 with a $(CA)_{23}$ dinucleotide repeat (Steffan et al. 1993) and HEL6 with $(GT)_{23}$ repeat (Kaukinen and Varvio 1993) were chosen for study as markers for paternity testing. The allele frequency in different breeds of cattle and the probability of exclusion of

incorrect sire in 15 breeds in the UK were determined in this thesis. Both CYP21 and ETH131 were included in the panel of markers used for the International Comparison Test in 1993.

The efficiency of parentage testing is measured in terms of the probability of excluding all but the true sire. Efficiency of parentage testing increases with the number of loci tested and the number and frequency of alleles at each locus. Different approaches have been used for calculating the probability of exclusion in animals (Gahne 1961, Jamieson 1965, Hayasaka et al. 1986, Weir 1990 and Jamieson 1994). All these methods have assumed Hardy-Weinberg Equilibrium (HWE) for the calculation. Techniques for genotype determination using VNTR markers in man have also assumed HWE (Devlin et al. 1990).

The artificial breeding practices have enabled extensive selection programmes to be carried out and thus HWE is not expected for cattle populations. With microsatellites, it is particularly important to account for sampling error where the number of alleles present at a locus is more than the number of individuals examined. In the present study the value of the five microsatellite loci has been examined for parentage verification, when used either independently or together, and their utility in several breeds of cattle has been assessed, taking into account deviations from HWE and linkage equilibrium and using appropriate statistical methods relevant to the small sample size.

3.2 MATERIALS AND METHODS.

3.2.1 *Animals*:

Two hundred and seventy five cattle belonging to 15 breeds in the UK were used in this study: 20 individuals in 10 breeds and 15 in another five. Samples were provided by the Cattle Blood Typing Service, Roslin. The pedigrees of the individuals included in the study were inspected and those chosen had no grandparents in common. The breeds used and the relevant abbreviations are listed in Table 3.2.

Table 3.2 Significance levels on testing for Hardy Weinberg Equilibrium

Breed		DRB3	CYP21	HEL6	ETH131	FSHB
Aberdeen Angus	(AA)	0.101	0.076	0.001	<0.001	0.014
Ayrshire	(AY)	0.059	0.309	<0.001	0.046	0.001
Belgian Blue	(BB)	0.001	0.083	0.074	0.009	0.449
Blonde Aquit.	(BA)	<0.001	0.114	0.148	0.331	<0.001
Dexter	(DX)	0.345	0.177	<0.001	0.488	<0.001
Charolais	(CH)	0.572	0.010	0.001	0.001	0.253
Friesian	(FR)	<0.001	0.004	0.013	0.146	0.162
Guernsey	(GU)	0.734	0.050	0.053	0.322	0.445
Hereford	(HF)	0.173	0.019	<0.001	0.052	0.368
Jersey	(JE)	<0.001	0.027	0.039	<0.001	<0.001
Limousin	(LN)	0.004	0.739	0.055	0.039	0.036
Simmental	(SG)	<0.001	0.002	0.004	0.004	<0.001
Belted Galloway	(BG)	0.459	0.024	0.105	0.039	0.013
South Devon	(SD)	0.175	0.005	0.001	0.432	0.642
Boran	(BO)	0.004	0.276	<0.001	<0.001	<0.001

Significance levels less than 0.05 indicate that the alleles may not be in Hardy Weinberg Equilibrium.

3.2.2 DNA preparation:

In cattle, peripheral blood leukocytes are the usual source of DNA for genotyping. The success of the technology is dependent on the separation of pure DNA from small quantities of blood. However, technical difficulties exist in the collection of blood samples from a large number of individuals in widely separated herds. Hence extraction of DNA from alternative sources was explored.

Milk samples with added preservatives can be stored up to 200 days without adverse effect on the polymerase chain reaction. Collection of milk samples is a part of daily routine and does not require trained personel. Semen samples can be stored in straws in liquid nitrogen for many years. Hair samples are readily available and do not require special preservation methods. The extraction of DNA by the simplest method from different tissues will be of great advantage in a field paternity testing service. Hence the procedures for extraction of DNA from these sources were explored and are given in chapter 2.

3.2.3 *PCR conditions:*

Regions containing microsatellite repeats were amplified by PCR. The primers used for amplification and the PCR conditions optimised for each locus are given in Table 3.3. The PCR was carried out as described in chapter 2. Radioactivity was incorporated either by adding α^{35}S dATP (Amersham) to the reaction mix or by using primers endlabelled with γ^{33}P ATP. The PCR protocol was denaturation at 94^{0}C for 1min. followed by annealing and extension at 60^{0}C for 5min for 30 cycles. For ETH131 annealing and extension was done at 55^{0}C for 5 min. The PCR products were stored at +4^{0}C.

A second round of PCR was done on those samples which failed to amplify in the first instance or when an insufficient PCR product was obtained. For this, 2ul of the PCR product from the first reaction was amplified in a second reaction under the same condition as for the first PCR.

The PCR products were electrophoresed on 6% denaturing polyacrylamide gels as described in chapter 2. The apparent sizes of alleles were assigned by comparison with a sequencing ladder from M13. Typing errors were minimized by

Table 3.3 Microsatellite loci used for parentage determination.

Locus	Dye	Primer sequence	PCR prog.	Mg conc.	Product size (bp)	Reference
DRB3	TAMRA	F GAGAGTTTCACTGTGCAG R CCAGAGTGAGTGAAGTATCTC	1	1.25mM	150-200	Ellegren et al. 1993
CYP21	JOE	F CCTGAAGAAGAGAATGGCAAC R GCTCGCGATCCAACTCCTCCTG	1	1.25mM	260-290	Chung et al. 1986
FSHB	FAM	F TGGGATATAGACTTAGTGGC R CAGTTTCTAAGGCTACATGGT	1	1.25mM	150-200	Kemp and Teale 1991
ETH131	HEX	F GTGGACTATAGACCATAAGGTC R GCTGTGATGGTCTACGAATGA	2	1.25mM	150-200	Steffan et al. 1993
HEL6	FAM	F GGACACGACTGAGCAAGTAA R AGGCAGATACATTACCACTA	1	1.5mM	260-290	Kaukinen and Varvio 1993

PCR programmes:

1. 30X (1'@94^{0}C, 5'@60^{0}C)

2. 30X (1'@94^{0}C, 5'@55^{0}C)

allele determination by two independent observers.

3.2.4 Statistical analysis.

To estimate the value of the microsatellite loci for parentage verification, allele frequencies, heterozygosity, PIC and the Probability that a RAndom Sire Excluded (PRASE) were calculated (see below). PRASE calculates the efficiency of markers for parentage testing and is a new method developed by Simpson et al (in prep).

With highly polymorphic markers, the number of individuals with each genotype is low and many combinations of alleles are not observed. A Monte Carlo method (Guo and Thompson 1992) was used to test for HWE and linkage equilibrium. The principle is as follows: the probability of observing the data under the assumption of HWE for a particular set of allele frequencies is calculated by random permutation to create sets of data that are typical of the results that might be obtained when the alleles are in HWE. To compare the original result with this, the proportion of the results with values smaller than the observed value is calculated. This proportion can be interpreted as a significance level which, if less than 5%, is deemed to show the locus is not in HWE.

Testing for linkage equilibrium is done both assuming HWE and non HWE, as appropriate for the locus. The techniques also depend on whether linkage disequilibrium is gametic phase disequilibrium or due to the excess of certain genotypic combinations. Because the parents in this study were not typed, gametic phase disequilibrium could not be tested directly.

Suppose the gametic phase is known and the loci are in linkage equilibrium, the frequency of the allele A_i is p_i and that of B_j is q_j. The estimation depends on whether the loci are in HWE or not. If the loci are in HWE the alleles at locus A and B are shuffled independently and two alleles from each locus are chosen to form new genotypes. If the loci are not in HWE, the genotypes A_iA_j and B_kB_l are shuffled and new genotypes are formed. This ensures that the genotype frequencies are the same as in the original sample. The probability of the new data is calculated and compared with the original value to estimate the significance.

The allele frequencies are determined for each microsatellite marker by direct gene counting method (Ceppelini et al. 1956, Smith 1957, Weir 1990).

3.2.4.1 *Heterozygosity.*

The level of heterozygosity also gives a measure of the usefulness of the marker. The markers with high percent of heterozygosity are more useful. Heterozygosity was calculated assuming either HWE or non HWE depending on the locus. Heterozygosity assuming HWE was calculated as:

$$1 - \Sigma p_i^2$$

where p_i is the frequency of i^{th} allele (Ott 1992).

3.2.4.2 *Polymorphic Information Content (PIC).*

The polymorphic information content is a measure of informativeness or usefulness of a marker for linkage studies. This was calculated using the formula

$$1 - \left[\sum_{i=1}^{n} p_i^2\right] - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2$$

where p_i and p_j are the population frequency of the i^{th} and j^{th} alleles (Botstein et al 1980).

Highly informative multi allelic systems such as VNTR or microsatellite loci therefore have PIC values close to one, whereas uninformative systems with low heterozygosity frequencies have PIC values close to zero.

3.2.4.3 *Probability of RAndom Sire being Excluded (PRASE).*

Probability of exclusion of an incorrect sire indicates the usefulness of a particular marker in paternity testing. To calculate PRASE all possible combinations of dam and sire genotypes were considered and the frequency of all the possible offspring genotypes were calculated. Each dam and offspring pair was considered

and the proportion of the sires in the population which can be excluded from parentage was calculated. The PRASE technique was developed as a part of this thesis (Usha et al 1995) and is explained in detail later in this chapter.

3.2.4.4 *Jackknifing.*

Jackknifing was used to determine the accuracy of the sample estimate. In this situation the PRASE was calculated for n samples containing n-1 animals, ie, deleting one observation each time. The variance between these values is $\frac{1}{n}\Sigma(\theta - \theta_i)^2$

where n is the original sample size, θ is the PRASE for the whole sample and θ_i is the PRASE for the sample of size (n-1) with the ith observation deleted. The variance of the sample of size n would be $\frac{n-1}{n}$ times this, therefore the Jackknifed SE is

$$\sqrt{\frac{n-1}{n^2}\Sigma(\theta-\theta_i)^2}$$

3.2.4.5 *Monte Carlo method.*

Because the observed number of alleles for microsatellites are more than the number of animals genotyped (ie. data with large number of alleles but a small sample size) a Monte Carlo method was used to test whether the observations were consistent with HWE and linkage equilibrium or not. This involves creating a new data set with the same size as that of the original by random permutation of the experimental data set.

3.3 RESULTS

DNA extracted from alternative sources - milk, semen, hair and tissues was compared to the DNA from blood for yield and purity. The amount of DNA was greater from blood and semen in comparison with other sources. Even though the

quantity of DNA per sample was less for milk and hair, the quality was sufficient for enzymatic amplification and genotyping for individual identification. The quality of DNA was tested by PCR amplification of the DNA from different sources using the primers flanking the DRB3 microsatellite loci. PCR products from blood, milk and semen were comparable and the bands on the gel were of equal intensity. The DNA extracted from tissues showed degradation. Poor amplification efficiency in DNA samples from tissues could be because of protein contamination.

The five markers evaluated for parentage verification in different breeds of cattle in the present study, DRB3, CYP21, ETH131, FSHB and HEL6, were selected on the basis of polymorphism, ease of typing and the possibility of multiplexing. Also DRB3 and CYP21 being linked, the question of efficiency of linked markers for parentage verification was addressed.

The five markers studied were found to be highly polymorphic in all the breeds of cattle studied. The pattern of bands generated after PCR amplification of each microsatellite locus on autoradiographs is shown in figure 3.1. The number of alleles, PIC and heterozygosity for each marker in each breed is given in Table 3.4. The DRB3 locus was very polymorphic having 23 alleles ranging in size from 144 to 220 base pairs. On average, each breed had ten alleles for this locus. Among the five loci studied, the DRB3 locus had the highest average PIC value of 0.93 with a heterozygosity of 0.65. The CYP21 microsatellite locus was also highly polymorphic in all the breeds, with 19 alleles varying in size from 254 to 300 base pairs. Taken together in all the breeds, the CYP21 locus had a PIC value of 0.91 and a heterozygosity of 0.78. DX and BG had fewer alleles (6 and 7 respectively) than FR and LN (13 and 12).

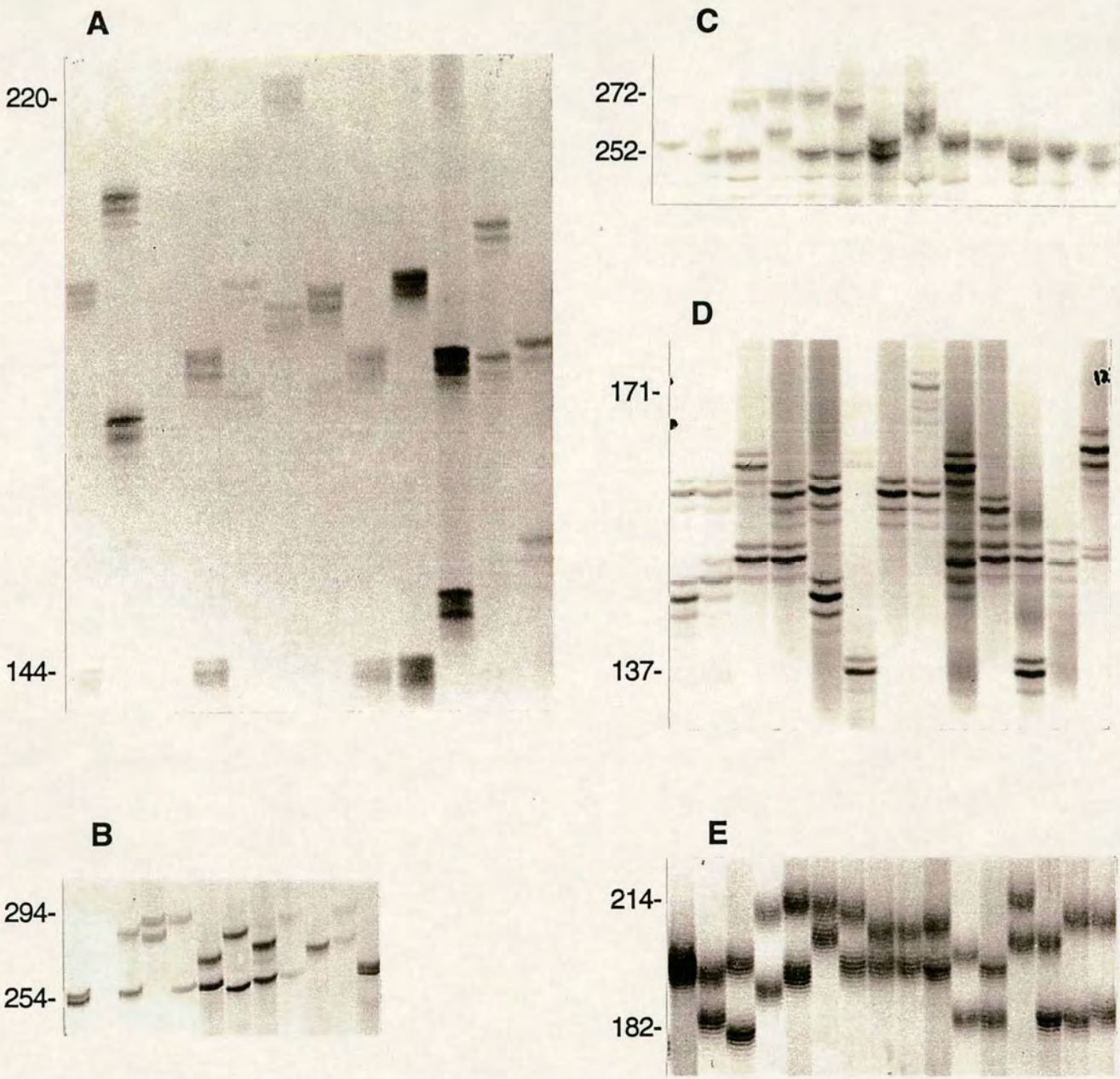
The HEL6 locus had 19 alleles in the population studied, with sizes ranging from 252 to 276 base pairs. Only two alleles were seen for BG, but twelve alleles were observed in AY. Twenty five alleles with sizes ranging from 137 to 171 base pairs were detected for ETH131. FSHB had 21 alleles with a size range of 182 to 222 base pairs. Excluding BG, the other breeds had on average nine alleles. These markers had PIC values of 0.88, 0.89 and 0.91 respectively.

For all the loci the alleles present in the different breeds and their frequencies

Table 3.4 Number of alleles, PIC and heterozygosity for the five microsatellite markers.

Breed	No:of Animals	MICROSATELLITE LOCI																							
		DRB3				CYP21				ETH131				FSHB				HEL6				MEAN ACROSS LOCI			
		Alleles	PIC	Heterozygosity		Alleles	PIC	Heterozygosity		Alleles	PIC	Heterozygosity		Alleles	PIC	Heterozygosity		Alleles	PIC	Heterozygosity		Alleles	PIC	Heterozygosity	
				HWE	Non-HWE			HWE	Non-HWE			HWE	Non-HWE			HWE	Non-HWE			HWE	Non-HWE			HWE	Non-HWE
AA	20	12	0.86	0.88	0.80	10	0.82	0.84	0.85	8	0.76	0.79	0.50	6	0.65	0.70	0.55	9	0.73	0.77	0.55	9.00	0.76	0.80	0.65
AY	20	9	0.83	0.85	0.70	10	0.83	0.82	0.80	9	0.75	0.78	0.70	11	0.78	0.81	0.45	12	0.84	0.85	0.50	10.20	0.81	0.82	0.63
BB	20	11	0.87	0.89	0.65	10	0.83	0.85	0.75	15	0.88	0.89	0.65	11	0.78	0.79	0.85	6	0.63	0.68	0.75	10.60	0.80	0.82	0.73
BA	15	8	0.68	0.70	0.40	10	0.83	0.84	0.80	10	0.80	0.82	0.80	8	0.76	0.79	0.40	9	0.76	0.79	0.73	9.00	0.77	0.79	0.63
CH	20	11	0.87	0.88	0.95	10	0.77	0.80	0.70	10	0.75	0.84	0.86	9	0.83	0.84	0.75	13	0.75	0.77	0.60	10.80	0.79	0.77	0.83
DX	15	7	0.73	0.77	0.66	6	0.69	0.73	1.00	7	0.83	0.78	0.65	6	0.68	0.72	0.60	5	0.68	0.72	0.20	6.20	0.72	0.62	0.74
FR	20	11	0.87	0.88	0.50	13	0.89	0.90	0.70	12	0.79	0.81	0.85	15	0.90	0.91	0.85	11	0.83	0.85	0.65	12.40	0.86	0.87	0.71
GU	20	11	0.82	0.84	0.90	9	0.82	0.84	0.75	10	0.73	0.77	0.80	13	0.68	0.70	0.75	10	0.80	0.82	0.75	10.60	0.77	0.79	0.79
HF	20	9	0.62	0.62	0.80	10	0.82	0.85	0.70	11	0.88	0.89	0.90	10	0.75	0.77	0.75	11	0.88	0.89	0.60	10.20	0.79	0.80	0.75
JE	20	11	0.85	0.87	0.60	10	0.85	0.87	0.65	6	0.72	0.75	0.65	11	0.87	0.88	0.45	7	0.65	0.70	0.65	9.00	0.79	0.81	0.60
LN	20	14	0.90	0.91	0.60	12	0.76	0.77	0.85	10	0.85	0.87	0.70	11	0.82	0.83	0.70	10	0.75	0.78	0.75	11.40	0.82	0.83	0.72
SG	20	11	0.75	0.78	0.50	10	0.86	0.87	0.70	10	0.77	0.79	0.70	11	0.86	0.87	0.55	8	0.78	0.80	0.65	10.00	0.79	0.82	0.62
BG	15	10	0.86	0.87	1.00	7	0.76	0.79	0.80	8	0.79	0.81	1.00	4	0.51	0.55	0.33	2	0.16	0.18	0.33	6.20	0.60	0.64	0.70
SD	15	7	0.77	0.80	0.93	10	0.85	0.86	0.73	6	0.75	0.78	0.86	7	0.76	0.79	1.00	7	0.66	0.69	0.33	7.40	0.69	0.78	0.84
BO	15	9	0.82	0.84	0.60	8	0.75	0.77	0.73	3	0.57	0.65	0.06	4	0.64	0.69	0.13	3	0.34	0.41	0.06	5.40	0.62	0.67	0.32

Figure 3.1 Banding pattern of five microsatellite loci



Autoradiograph showing the PCR products of five microsatellite loci labelled with ^{35}SdATP. Alleles in base pairs are displayed on the left of each group. a) DRB3 b) CYP21 c) HEL6 d) ETH131 e)FSHB

generally varied considerably, eg: alleles present at high frequency in one breed were absent in another. For instance, an allele of the DRB3 locus with a frequency of 0.6 in HF was absent in JE and FR. The frequency of alleles for different markers against each breed are given in Tables 3.5 to 3.9. A large number of CYP21 alleles were present in the FR and LN, covering a wider size range, whereas fewer alleles present in DX, BG and BO which covered a smaller size range. Similarly four alleles each were detected in BG and BO in contrast to 15 alleles in FR for FSHB. With HEL6, twelve alleles were present in AY compared with two in BG and three in BO. In general DX, BG and BO had fewer alleles for all the five loci when compared with the other breeds. Average heterozygosity across loci within each breed is shown in table 3.4.

The breeds in this study deviated from HWE by varying degrees. AA, AY, GU and DX were found to be in HWE for the CYP21 and DRB3, whilst FR, JE and SG were in equilibrium for none of the markers. When all the five loci were considered together JE and SG exhibited deviations from HWE ($p < 0.05$) whereas GU did not exhibit deviations for any of the loci ($p > 0.05$). The remaining breeds showed significant deviations from HWE at some of the loci ($p < 0.05$). Significant levels are shown in Table 3.2. Generally, in all the breeds studied, the loci for which there was an excess of homozygotes, deviated from HWE. For example, in BA out of the 15 animals, 9 were homozygous for DRB3. Similarly, in AY 11 animals were homozygous for FSHB ie. in cases where the number of homozygotes exceeded 50% the breeds deviated from HWE.

As DRB3 and CYP21 are linked, statistical analysis was done to verify whether they were in linkage equilibrium in the population. Only AA, AY and BG exhibited significant deviations from linkage equilibrium for these loci at the population level (Table 3.10).

The PRASE technique was developed (in conjunction with Dr S.P. Simpson) as a part of the work in this thesis and is used to measure how useful a marker is for parentage testing. It determines the proportion of sires in a population which can be excluded from paternity using a particular marker for a randomly selected dam-calf pair. PRASE is calculated from the allele frequencies of loci in the population

Table 3. 5 No: of alleles for the marker CYP21 in 15 breeds of cattle.

	AA	AY	BB	BA	DX	CH	FR	GU	HF	JE	LN	SG	BG	SD	BO	TOTAL
254	0	1	0	1	0	0	4	0	1	0	3	0	0	0	0	10
258	0	7	0	0	0	0	0	0	4	4	0	2	7	0	0	24
260	8	8	5	3	3	2	1	10	7	0	2	4	10	7	3	73
262	3	1	5	0	3	0	0	0	3	0	0	6	3	5	5	34
264	0	12	10	0	2	2	1	1	10	2	3	8	4	4	3	62
266	5	2	5	0	13	3	3	6	7	5	0	4	4	3	1	61
268	1	2	0	9	7	9	0	4	0	6	2	2	1	3	12	58
270	8	0	1	2	2	2	3	6	1	4	18	4	1	1	0	53
272	2	0	3	2	0	14	6	3	0	8	2	3	0	1	1	45
274	9	2	1	4	0	5	2	7	1	7	1	6	0	2	0	47
276	1	4	7	2	0	0	2	0	2	2	0	1	0	0	0	21
278	2	0	0	4	0	0	6	2	4	1	1	0	0	1	1	22
280	0	1	1	2	0	1	1	1	0	0	0	0	0	0	0	7
282	0	0	2	1	0	1	3	0	0	0	0	0	0	0	4	11
284	1	0	0	0	0	1	5	0	0	1	0	0	0	0	0	8
288	0	0	0	0	0	0	3	0	0	0	2	0	0	3	0	8
290	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2
294	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	3
300	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1

Table 3.6 Number of alleles for the marker DRB3 in 15 breeds of cattle

	AA	AY	BB	BA	DX	CH	FR	GU	HF	JE	LN	SG	BG	SD	BO	TOTAL
144	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4
152	0	0	0	0	8	8	0	0	0	2	0	0	0	1	0	19
154	0	0	2	0	0	2	0	2	0	3	0	0	0	0	0	9
156	4	3	6	1	10	0	0	1	1	10	1	3	0	0	0	40
160	1	0	0	0	0	0	0	0	0	2	0	0	2	0	0	5
164	0	0	2	0	0	0	0	0	0	0	2	0	0	0	3	7
166	0	0	0	1	0	0	0	5	0	0	0	0	0	5	8	19
168	4	0	0	0	0	0	3	5	0	5	0	0	2	0	0	19
170	2	0	4	0	0	0	1	6	0	4	1	0	0	4	1	23
172	0	0	0	0	0	1	1	3	2	1	2	1	3	0	1	15
174	7	0	4	0	4	7	4	1	0	2	3	1	2	0	2	37
176	1	10	6	0	5	4	5	1	4	6	4	1	0	0	2	49
178	1	4	3	0	1	2	6	0	2	2	4	1	5	3	6	40
180	2	4	6	2	0	3	6	12	2	3	3	2	0	0	0	45
182	1	0	3	0	1	3	0	0	2	0	4	0	3	0	0	17
184	7	3	3	15	0	2	5	1	2	0	5	8	1	1	0	53
186	0	7	0	4	0	4	2	3	24	0	6	16	1	0	5	72
188	0	6	0	4	0	0	1	0	1	0	2	0	0	8	2	24
190	7	0	0	0	0	4	0	0	0	0	0	0	5	0	0	16
194	0	0	0	2	1	0	0	0	0	0	0	2	0	0	0	5
200	0	2	0	0	0	0	6	0	0	0	2	1	6	8	0	25
214	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
220	0	0	1	0	0	0	0	0	0	0	1	4	0	0	0	6

Table 3. 7 Number of alleles for the marker ETH 131 in 15 breeds of cattle.

	AA	AY	BB	BA	DX	CH	FR	GU	HF	JE	LN	SG	BG	SD	BO	TOTAL
137	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
138	11	0	2	0	0	0	0	0	0	0	0	1	0	9	0	23
141	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	4
143	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10
145	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	3
146	0	0	0	0	0	0	1	0	0	0	6	0	0	0	13	20
148	0	6	1	0	5	2	0	1	5	6	0	1	0	0	0	27
149	6	14	5	8	11	11	12	12	4	6	7	6	0	1	7	110
151	12	10	6	4	0	5	11	11	3	7	3	9	5	7	0	93
152	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	9
153	0	1	1	1	0	1	1	0	7	1	0	0	1	6	0	20
154	0	0	3	1	0	0	2	0	0	0	0	2	0	0	0	8
155	5	0	3	1	0	2	2	0	0	0	2	1	1	0	0	17
156	0	2	9	2	3	0	4	1	3	4	4	0	0	0	0	32
157	0	2	1	8	0	3	0	10	0	16	9	3	4	0	0	56
158	0	0	1	2	0	0	1	0	0	0	2	14	0	0	0	20
159	0	2	0	2	1	6	1	1	5	0	2	2	0	2	0	24
160	2	2	1	0	4	0	0	0	1	0	2	0	0	0	0	12
161	1	0	2	0	0	6	2	1	1	0	0	1	0	0	0	14
162	0	1	1	0	0	0	0	0	0	0	3	0	0	0	0	5
163	1	0	0	0	0	1	0	0	4	0	0	0	3	5	0	14
164	2	0	0	0	0	0	0	0	0	0	0	0	6	0	0	8
167	0	0	0	1	1	3	0	0	0	0	0	0	0	0	0	5
169	0	0	0	0	0	0	0	1	3	0	0	0	1	0	0	5
171	0	0	0	0	5	0	0	0	4	0	0	0	0	0	0	9

Table 3.8 Number of alleles for the marker FSHB in 15 breeds of cattle.

	AA	AY	BB	BA	DX	CH	FR	GU	HF	JE	LN	SG	BG	SD	BO	TOTAL
182	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
184	0	0	1	0	0	0	2	1	0	0	0	0	0	0	0	4
186	0	0	0	2	5	0	0	0	0	4	0	0	0	0	0	11
188	0	0	0	0	0	2	1	0	2	4	0	0	0	0	0	9
190	0	0	16	3	2	15	2	21	6	5	3	6	0	0	0	79
192	5	2	1	0	0	1	2	1	1	0	2	5	0	10	0	30
194	18	7	4	0	0	0	2	4	0	0	2	0	0	0	6	43
196	0	0	0	0	1	0	3	2	1	1	0	1	2	0	14	25
198	0	2	3	6	7	0	1	0	17	0	8	1	0	4	5	54
200	0	0	2	10	0	2	7	2	0	2	5	4	19	5	0	58
202	3	0	4	1	2	12	6	2	2	0	12	8	4	0	0	56
204	11	11	3	6	0	0	3	0	0	8	1	0	0	1	0	44
206	1	11	1	1	1	0	0	2	2	0	1	6	5	0	0	31
208	0	0	1	0	4	8	0	1	0	0	3	0	0	0	5	22
210	0	2	4	0	2	0	1	1	2	2	0	1	0	6	0	21
212	0	0	0	1	0	0	2	2	0	2	1	4	0	1	0	13
214	2	0	0	0	6	0	4	1	4	2	0	3	0	3	0	25
216	0	2	0	0	0	0	3	0	0	5	0	0	0	0	0	10
218	0	2	0	0	0	0	0	0	3	5	2	0	0	0	0	12
220	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
222	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 3.9 Number of alleles for the marker HEL 6 in 15 breeds of cattle.

	AA	AY	BB	BA	DX	CH	FR	GU	HF	JE	LN	SG	BG	SD	BO	TOTAL
252	0	0	0	0	0	0	0	1	0	0	0	0	0	0	22	23
256	0	0	0	0	0	0	3	0	2	2	7	0	0	0	0	14
257	0	1	0	0	13	4	3	0	4	2	0	0	0	0	0	27
258	1	2	19	3	5	5	8	3	7	13	10	9	0	15	0	100
259	12	6	0	0	0	5	0	6	4	0	0	0	27	0	0	60
260	4	5	2	1	2	0	1	0	5	0	0	0	0	0	0	20
261	14	11	0	1	6	0	0	3	5	0	0	0	0	3	0	43
262	2	4	8	1	0	16	3	1	4	1	0	13	0	1	7	61
263	3	0	1	0	0	0	0	4	0	1	1	0	0	0	0	10
264	0	5	0	0	0	0	4	14	2	17	5	3	0	2	0	52
265	0	2	0	0	0	0	0	3	2	0	1	0	0	0	0	8
266	2	1	9	9	0	0	3	0	0	0	1	2	0	0	0	27
269	0	0	0	0	0	1	0	4	2	0	0	4	0	0	0	11
271	1	1	1	9	4	0	11	0	0	4	13	4	3	6	1	58
272	0	1	0	2	0	2	2	0	0	0	1	0	0	2	0	10
273	0	0	0	1	0	6	1	0	0	0	0	3	0	0	0	11
274	0	0	0	3	0	0	0	0	0	0	0	0	0	1	0	4
275	0	1	0	0	0	1	1	1	3	0	1	0	0	0	0	8
276	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3

Table 3.10 Significance levels on testing for Linkage Equilibrium between the markers CYP21 and DRB3

Breed	Hardy-Weinberg equilibrium	Non Hardy-Weinberg equilibrium
AA	0.012	0.025
AY	<0.001	0.002
BB	0.056	0.748
BA	0.003	0.361
CH	0.261	0.315
DX	0.824	0.828
FR	<0.001	0.780
GU	0.338	0.460
HF	0.021	0.293
JE	<0.001	0.262
LN	0.189	0.829
SG	0.007	0.550
BG	0.020	0.004
SD	0.472	0.399
BO	0.005	0.117

Significance levels less than 0.05 indicate the loci may be in linkage disequilibrium.

from which the dams, putative sires and the true sire were selected. Each dam-calf pair was considered and the proportion of the sires in the population which can be excluded from the parentage was calculated. For a population in HWE and for a single locus, the method of Chakroborty et al (1988) can be used. This method was extended in this thesis to account for deviations from HWE, for analysis of multiple loci, and to verify whether the loci are in linkage equlibrium. This is achieved by using observed genotypes instead of allele frequencies. In its most general form, calculation of PRASE can be expressed as

$$\text{PRASE} = \sum_{d} \sum_{s} P(G_d) P(G_s) \sum_{o} P(G_o \mid G_d, G_s) \sum_{p} P(G_p) P(\text{excluded} \mid G_d, G_p, G_o)$$

where $P(G_d)$, $P(G_s)$, $P(G_p)$ are the genotype frequencies of the dams, true sires and the putative sires. The term $P(G_o/G_d,G_s)$ is the transmission probability, ie the proportion of offspring with genotype G_o when the parental types are G_d and G_s. The final term $P(\text{excluded}/G_d, G_p, G_o)$ is one or zero depending on whether the putative parental genotypes, G_d and G_p, could have produced the genotype G_o. For two loci in linkage equilibrium PRASE is $\text{PRASE}_{AB} = 1-(1-\text{PRASE}_A)(1-\text{PRASE}_B)$.

In practice, PRASE is calculated by enumerating all possible offspring genotypes for each pair of parental genotypes, then listing and checking which of the putative parental genotypes could be excluded. These are then weighted by their frequencies and summed to determine the PRASE. For example, for a single locus with two alleles, A and a, if G_d and G_s are AA (ie the true parents are both AA) the offspring will be AA with a probability of 1. When G_p is AA or Aa the putative sire *could* be the true sire, so $P(\text{excluded} \mid G_d, G_p, G_o) = 0$, however if G_p is aa the putative sire *cannot* be the true sire and $P(\text{excluded}/G_d, G_p, G_o) = 1$. PRASE is found by summing over all possible genotype combinations. The technique of jacknifing was used to measure the accuracy of the sample estimate. Since the sample size used in the analysis was small, the technique of bootstrapping was used to generate new data sets.

PRASE indicates the value of a marker for parentage verification. Calculation of PRASE is dependent on whether or not the loci were found to be in

HWE and in linkage equilibrium or disequilibrium, and is calculated on the basis of allele frequencies in the population of dams, putative sires and the true sire. The results for PRASE calculated for the 15 breeds are shown in Table 3.11. PRASE values ranged from 0.44 to 0.85 and 0.41 to 0.81 for DRB3 and CYP21 loci respectively within breeds. The highest PRASE value was for LN & FR and the lowest was for HF & DX for DRB3 and CYP21 respectively. Used independently DRB3 and CYP21 had a PRASE of 0.72 and 0.68 respectively taking the data as a whole. For HF, BA, GU, SG and SD the CYP21 locus had a higher PRASE than for DRB3. For other breeds DRB3 locus had a higher PRASE value. Taking them together, both markers would give a PRASE of 0.88 for all the breeds, with a value ranging from about 0.78 in the DX to better than 0.94 in the FR.

Thus using CYP21 and DRB3 markers in the cattle population as a whole it would be expected that in 88% of cases an incorrect sire would be identified. These markers are however closely linked (1.4cM). If unlinked markers with similar frequencies were used the value would be 91% (Table 3.12). This shows that the use of tightly linked markers do not reduce the PRASE value considerably if the markers are equally polymorphic and are in linkage equilibrium in the population studied.

ETH131 had PRASE values ranging from 0.56 to 0.99 in the breeds studied. Among the five markers used HEL6 had the lowest PRASE value of 0.15 for BG with values ranging between 0.15 to 0.81. For FSHB the lowest value was 0.41 for BG and highest of 0.83 for JE. Pooled data across breeds for ETH131 gave a PRASE value of 0.66 while HEL6 and FSHB had values of 0.61 and 0.65 respectively when used independently. However, the value of the markers differed considerably between breeds (Table 3.11).

3.4 DISCUSSION

The purpose of this study was to evaluate the precision of microsatellite markers for parentage verification in different breeds of cattle. The high polymorphism, relative abundance and amenability of the markers to assay by PCR amplification explain their general acceptance for linkage studies and individual

Table 3.11 Calculation of Probability of Random Sire Excluded (PRASE)

Breed	DRB3	CYP21	ETH131	HEL6	FSHB	Total
AA	0.773	0.688	0.711	0.660	0.526	0.996
AY	0.751	0.671	0.585	0.788	0.718	0.997
BB	0.820	0.730	0.804	0.416	0.595	0.997
BA	0.668	0.685	0.649	0.641	0.700	0.996
CH	0.774	0.621	0.715	0.695	0.490	0.996
DX	0.611	0.409	0.574	0.595	0.775	0.992
FR	0.831	0.812	0.661	0.753	0.810	0.999
GU	0.779	0.722	0.561	0.696	0.509	0.993
HF	0.441	0.748	0.775	0.814	0.628	0.997
JE	0.797	0.756	0.609	0.475	0.829	0.998
LN	0.852	0.586	0.765	0.612	0.730	0.998
SG	0.695	0.775	0.626	0.694	0.808	0.998
BG	0.720	0.582	0.604	0.146	0.405	0.978
SD	0.579	0.752	0.562	0.614	0.536	0.992
BO	0.732	0.602	0.618	0.398	0.635	0.991

Table 3.12 Probability of Random Sire Excluded(PRASE): DRB3 and CYP21.

Breed	DRB3	CYP21	Linked	Unlinked
AA	0.773	0.688	0.887	0.929
AY	0.751	0.671	0.874	0.918
BB	0.820	0.730	0.924	0.952
BA	0.668	0.685	0.862	0.895
CH	0.774	0.621	0.875	0.903
DX	0.611	0.409	0.775	0.770
FR	0.831	0.812	0.937	0.968
GU	0.679	0.722	0.888	0.911
HF	0.441	0.748	0.843	0.859
JE	0.797	0.756	0.917	0.950
LN	0.852	0.586	0.916	0.930
SG	0.695	0.775	0.907	0.931
BG	0.720	0.582	0.823	0.883
SD	0.579	0.752	0.861	0.896
BO	0.732	0.602	0.852	0.893
Overall	0.719	0.682	0.880	0.910

identification (Tautz 1989). Microsatellite markers are considered superior to other markers, because they require small quantities of DNA for analysis and are amenable to automation. Microsatellites provide a relatively new tool in the field of parentage verification because these type of markers overcome the disadvantages of other systems - blood typing, protein polymorphism, restriction fragment length polymorphism and minisatellites.

Before a microsatellite locus is accepted for parentage verification it must, like other genetic markers, fullfil several basic criteria such as Mendelian inheritance, location on different chromosome, absence of interaction with environmental factors, and a high PIC and heterozygosity value. Furthermore, routine parenthood analysis requires narrow allele size ranges, simple and reproducible detection of the alleles and the possibility of multiplexing. The use of microsatellites for paternity testing has become popular. Perkin Elmer is marketing a "StockMarks" kit which permits automated genotyping of cattle and contains a set of 11 markers.

For routine work it is essential to have a procedure which quickly and effectively releases DNA from blood and semen samples. In addition to blood and semen samples, milk and hair formed a good alternative source for DNA. The technical convenience of milk as a source of DNA increases the field of application in parentage testing. Semen from bulls being stored for a long time enable them to be genotyped after many years. Though alternative source of DNA from easily accessible tissues like hoof, horn or even from urine can be tried, these methods are not in common use. The amount of DNA will be poor when compared to blood and for urine the chance of contamination is increased. The method is not required to yield high molecular weight DNA, because the PCR technique amplifies small target sequences of DNA. The extraction procedures should also be rapid and require only a few steps of manipulation in order to minimize the possibility of cross-transfer and cross-contamination.

The precision of allele size designation across the gels is of paramount importance for routine parentage verification work to enable comparison of results between gels. In this study, for the loci, ETH131 and FSHB, some alleles varied

in size by just one base pair. In the sequence for FSHB there is a stretch of $(T)_{13}$ adjacent to the dinucleotide repeat which may be responsible for single base pair differences in size (Fries et al. 1990). Slippage of Taq DNA polymerase during amplification may also be a problem. According to Glowatski Mullis et al. (1995), the sequence of two fragments of ETH131 microsatellite which has the same overall length in two different runs differed in the number of CG and CA dinucleotide repeats. The difference in the number of the dinucleotide repeats caused a mobility shift by one base pair because CG repeats are sensitive to slight variation in the denaturing capacity of different gels.

A disadvantage of dinucleotide repeat polymorphism is that each allele consists of not just one fragment, but a series of fragments. These shadow bands obscure the position of allelic fragments, which can make genotyping difficult or impossible when typing individuals heterozygous for alleles differing in length by only two nucleotides (Murray et al. 1993, Hauge and Litt 1993). There is a need to standardize the sizing and numbering of microsatellites as well as the subsequent estimation of allele frequency so that maximum benefit can be derived from these highly polymorphic markers.

Five different factors are known to cause shadow banding. These are: 1. internal labelling of PCR products, 2. amplification of nontarget regions of the genome to produce products that interfere with visualization of desired products, 3. insufficient denaturing conditions during gel electrophoresis, 4. the partial nontemplate dependent addition of a single nucleotide to the 3' end of a PCR product by Taq polymerase (Weber 1989), and 5. the insertion or deletion of repeat units during the PCR by slipped-strand mispairing to produce a two nucleotide ladder of fragments (Weber and May 1989, Luty et al. 1990, Hauge and Litt 1993). The problem with internal labelling of PCR products arises because of the separation of the CA and GT strands in denaturing gels (Weber and May 1989). When the PCR products are labelled internally with a single radioactive nucleotide and the two strands differ sufficiently in their content of that nucleotide, strand separation causes less interference with the scoring of alleles. Both strands are labelled with equal intensity if the base composition between the strands is similar, which makes scoring

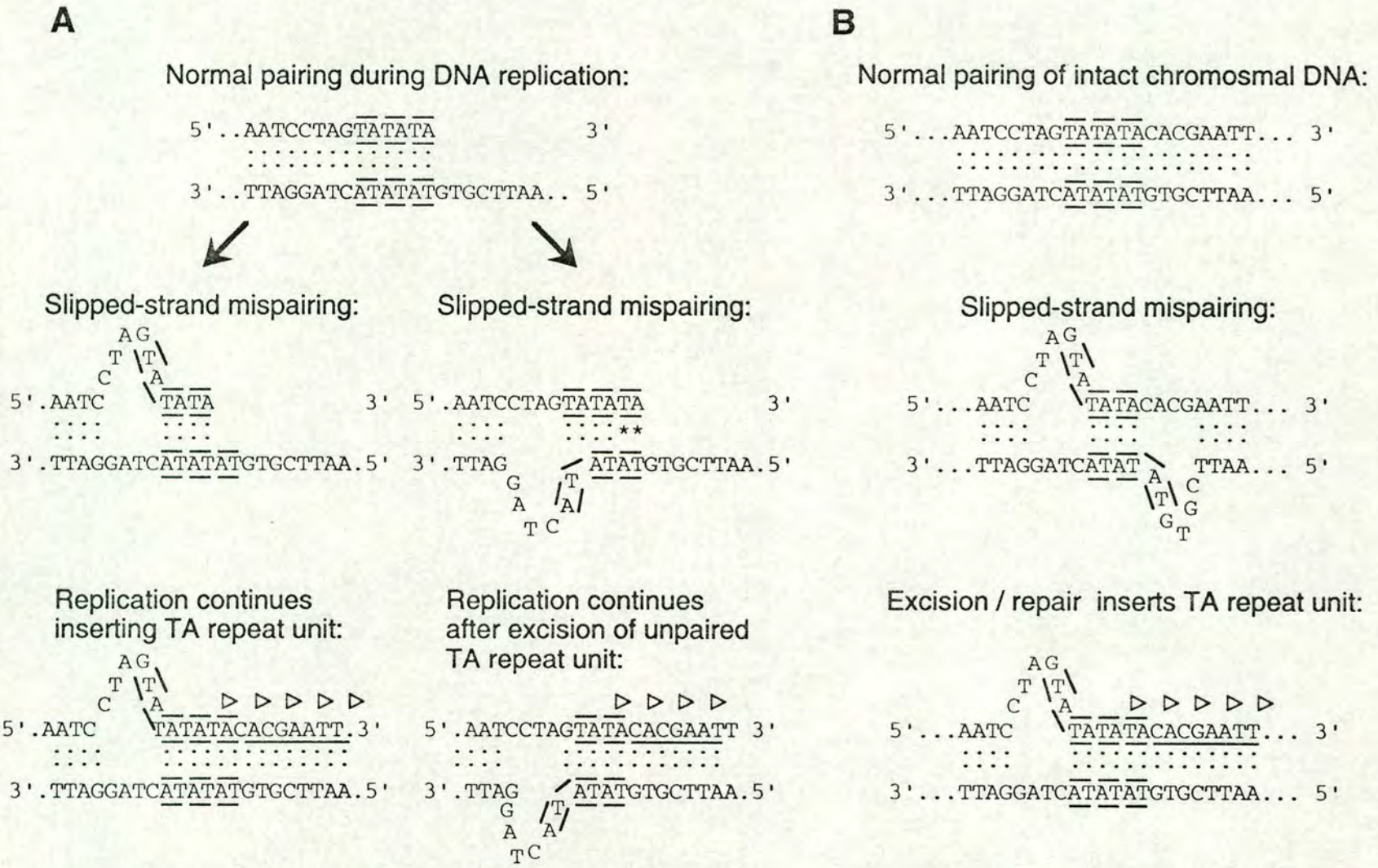
of alleles difficult. This problem is avoided by labelling only one of the PCR primers at its 5' end with γ^{33}P ATP so that only one of the two strands is visualized by autoradiography (Weber and May 1989) resulting in a simpler pattern.

Amplification of the nontarget regions of the genome varies greatly between different marker loci and is usually minimized by appropriate design of the primers and by choosing sufficiently stringent annealing conditions. The use of "touch down" PCR protocols (Don et al. 1991) helps to minimize the problem of nontarget annealing by permitting the formation of only perfectly matched primer template. Even though the use of formamide in the gels improves the denaturing conditions to enhance the quality of typing with dinucleotide repeats (Litt et al. 1993), the banding pattern of FSHB and ETH131 used in this study could not be improved.

Slipped-strand mispairing can occur during the process of chain elongation by *Taq* DNA polymerase where a region of the primer - template complex becomes unpaired during primer extension allowing slippage of either primer or template strand such that one or more repeat units form a non base-paired loop. As a consequence, the primer strand becomes lengthened or shortened by a length equal to an integral multiple of the length of the repeating unit (fig 3.2). The chances for the template strand to loop are greater because it is longer than the strand being elongated. Shorter shadow bands implies that the looping out of the template is a main mechanism for the production of shadow bands. In the process, termed *recombination*, partially extended PCR products can hybridise to another template which are away from the CA repeat regions (Murray et al. 1993). If these products are extended, they will be of a different size from the main band. This has serious implications for multiplexed PCRs since extra products complicate the analysis of products. Shadow bands do not occur with tetranucleotide repeats (Edwards et al. 1991).

Another disadvantage with microsatellites is the non-amplifying alleles (Pemberton et al. 1995). These alleles are detected when a mutation is found in one or more of the priming sites (Koorey et al. 1993, Callen et al. 1993). Null alleles often result in apparent homozygote excess. In theory, non-amplifying alleles can be revealed either by lowering the priming stringency (to accomodate sequence

Figure 3.2 Single-stranded mispairing



Generation of duplications or deletions by SSM between contiguous repeats. Small arrows indicate direction and initiation of DNA synthesis: colons indicate base pairing. A) 2-base slippage in an AT- repeat during replication of a DNA duplex, followed by continued chain elongation. Slippage in the 3'-5' direction (left panel) results in insertion of one AT unit; slippage in the other direction (right panel) results in deletion of one repeat unit. B) similar slippages occurring in intact duplex DNA. Mismatched regions form single stranded loops, excision of the shorter loop on the top strand, followed by repair synthesis using the lower strand as the template results in addition of one AT repeat unit.

mismatches) or redesigning the primers (to avoid mutation sites).

The markers chosen in the present study were the ones that were found highly polymorphic from the microsatellite markers available at the time. A panel of 25 microsatellites were tested on 20 Friesian DNA samples and the markers were chosen based on polymorphism, ease in amplification and allele designation.

The five markers used in this study were highly polymorphic and informative. Two of the markers were linked and the other three were located on different chromosomes. In the study by Ellegren et al. (1993) involving three breeds of cattle, 14 alleles were described for DRB3, as compared to 23 alleles in 15 breeds in this study. Similarly 13 alleles for CYP21 were found in the International genome mapping reference panel (Fries et al. 1990), but 19 alleles in the present study. Eleven alleles were detected for FSHB in the CSIRO cattle reference families (Moore et al. 1992) as against 21 alleles in the present study. Likewise, in the four breeds studied earlier, only eight alleles were recognised for ETH131 (Steffan et al. 1993), and ten alleles for HEL6 in Finnish cattle (Kaukinen and Varvio 1993). In contrast, 25 alleles were detected for ETH131 and 19 alleles for HEL6 in the current study. It is possible that the greater number of alleles recognised in this study was due to the use of unrelated individuals from 14 breeds, thus giving a broader genetic base than used in the previous studies. Clearly, all these markers have high levels of polymorphism that contributed to the efficiency in parentage verification in most populations.

It is important to have a measure that a particular sire is taken as the true sire in parentage investigation. This depends on the frequency of alleles of the microsatellite loci in the population. If an allele is common in the population the probability of a coincidental matching is high, and vice versa. To obtain the frequency of a combination of paternal alleles an estimation of the allele frequencies in the general population is required. The precision of the estimate increases with increased sample size. In the present study the sample size is small, so the allele frequencies may not reflect their values in the whole population.

The JE and FR had uniformly distributed allele frequencies with no allele being present at a much higher frequency than others for DRB3 and CYP21 loci.

But in BG, fewer alleles were seen for all the five markers. This restricted range of allele sizes may reflect expansion of the breed following a genetic bottleneck. The smaller number of alleles seen in these breeds could be the result of a smaller number of sires used or through inbreeding. The smaller sample size might also have an effect in the number of alleles.

In man, HWE is assumed for calculating the probability of exclusion in paternity cases. In the earlier described methods of calculation of probability of exclusion in cattle, HWE was assumed (Gahne et al. 1961, Jameison 1965, Hayasaka et al. 1986, Weir 1990). In livestock populations, because of extensive artificial insemination practices it is necessary to check if the population is in equilibrium. To obviate this situation, the method described by Chakraborty et al. (1988) has been modified in this study to take into account HWE and linkage equilibrium. For highly polymorphic markers like the microsatellites, the number of individuals with each genotype was low and many combination of alleles were not observed. The Monte Carlo method of Guo and Thompson (1992) was used to estimate the exact significant levels of the tests for HWE for multiple alleles.

Only four breeds, AA, AY, DX and GU, were found to be in Hardy-Weinberg equilibrium for DRB3 and CYP21 loci. In these breeds, selection, either natural or for production traits, appears not to have distorted the population when these two loci were considered. For most breeds, the use of artificial insemination has enabled extensive selection programmes to be carried out, and the selection of an extreme phenotype might be expected to produce disequilibrium. It is therefore, not surprising to find that for JE and SG neither CYP21 or DRB3 locus was in equilibrium, which may reflect selection for a phenotype affected by these genes. The MHC DR antigen is involved in antigen presentation, and thus influences immunity to particular diseases, whilst CYP21 is involved in steroid metabolism and may affect production traits. When all the loci were considered together only GU was found to be in HWE for all the loci while JE and SG were not in equilibrium for any of the five loci studied. Apart from the factors mentioned above, the sampling strategy could have an influence on HWE. Also the admixture of different sub populations within a breed can explain the distortion of equilibrium in the

population.

Despite these loci being tightly linked, at 1.4cM on the latest bovine map (Barendse et al. 1994), DRB3 and CYP21 loci do not appear to be in linkage disequilibrium. Thus the alleles present at one locus do not predict those found at the other, at least at the population level. This is important when using these markers for paternity verification. PRASE gives a measure of the usefulness of a marker for parentage verification. It is seen that the use of tightly linked markers DRB3 and CYP21 does not appreciably reduce the PRASE when compared with similar unlinked markers (Table 3.12). Within a single population, linkage disequilibrium tends to reduce PRASE (Simpson et al. in prep). An accurate PRASE value can be calculated only when both the nature and magnitude of linkage disequlibrium is known. The use of both DRB3 and CYP21 significantly increases PRASE as compared with using either one or the other alone. Therefore, unless there is strong linkage disequilibrium between two loci, a highly polymorphic second linked locus is preferable for paternity testing to a less polymorphic unlinked locus.

Among the five markers used, DRB3 gave the highest PRASE value of 0.85 in LN and the lowest value of only 0.15 in BG with HEL6. For BG, the highest PRASE value was 0.72 with DRB3. This clearly indicated that HEL6 is not a good marker for parentage determination in BG compared with DRB3. But in BA all the five markers gave a similar PRASE value. The value of the markers differed between breeds.

The PRASE calculated is accurate for this data set, but the data are only a sample for the whole population. To measure the variance of probability, standard errors were calculated using the technique of Jackknifing. The PRASE calculated depended upon whether the sire and the dam were drawn from a single homogenous population or from a population differing in allele frequency. When the dam and the true sire belonged to different populations, the effect on the PRASE was minor compared to the situation when the true sire and putative sire belonged to different populations (Simpson et al. in prep). The best loci for parentage testing within a breed may not be the best loci where the breed of the putative sire is unknown or when crossbred cattle is used.

The calculation of paternal exclusion probabilities depends on the estimation of allele frequencies in the population from which the putative sire is drawn. In the present study, when a single locus was considered, the exclusion probability was generally lower with small sample size. This could be because for multialleleic systems like microsatellites, it is possible that some alleles may not be observed in smaller samples which will affect the calculation of probability of exclusion. But when data on several loci are included the reduction in exclusion probability due to small sample size is diminished (Hill pers comm.).

Previously reported methods of estimating PRASE have assumed Hardy Weinberg and linkage equilibrium (Chakraborty et al 1988). On the basis of the data from this study it is seen that most of the breeds are not in equilibrium. In most cases, assumption of equilibrium overestimated the power of markers, although this may not always be the case. In BG and BO there was an excess of homozygotes in the test panel and for these breeds a lower PRASE value was obtained assuming Hardy Weinberg Equilibrium. The best estimate of PRASE is obtained when an accurate estimate of the population genetic structure is taken into account. The "StockMarks" kit marketed by Perkin Elmer uses 11 markers which are less polymorphic. The number of alleles varies between three and nine for these markers and the heterozygosity of 0.46 - 0.84 based on the calculation using 17 Holstein sires. The probability of exclusion of incorrect sire is low for each marker, but with 11 markers the probability of exclusion is estimated to be 0.999. In the present study, because the markers were highly polymorphic, with 5 markers, greater than 99% probability of exclusion was achieved for most of the breeds. Moreover, as the data with the "StockMarks" kit is based on one breed, the probability may change when used on different breeds. This clearly indicates that the polymorphic nature of the markers is inversely proportional to the number of markers to be used for parentage verification to achieve maximum probability of exclusion.

Used together the CYP21 and DRB3 would generally give 88% probability of excluding an incorrect sire. The inclusion of a third locus FSHB, with a similar level of polymorphism increased the value to about 97% which is comparable with the estimated efficiency of parentage verification using conventional serological

techniques at the Cattle Blood Typing service (Spooner and Ross pers comm.). Addition of two more markers - HEL6 and ETH131 increased the probability of exclusion to better than 99% in all the breeds studied. Typing of such microsatellite loci could be achieved on one track of an electrophoretic gel, by multiplexing which could be of great value in a regular typing service. The use of multiplexing is described in detail in chapter 4.

In conclusion, microsatellite loci have attractive properties and if fully exploited, will replace conventional blood typing for parentage varification in cattle in the near future. They are not only highly informative but also relatively simple to use and standardize. Using multiplex PCR procedures and automated detection techniques, microsatellite typing at present appears to be the simplest way to clarify family relationships and to provide genetic evidence of parentage. Although the efficiency of the microsatellite loci in this study is very high, they may still not represent the best combination of markers for parentage testing. On the contrary, the results presented in this study might stimulate the investigation of other microsatellite loci for their applicability to parentage analysis. A panel of markers has to be set which gives the highest probability of exclusion for all the breeds. The PCR based technology will replace the conventional blood typing methods for parentage determination within the next decade. The use of well tested markers will be an important criterion in this context.

CHAPTER 4

DEVELOPMENT OF METHOD FOR ROUTINE TYPING OF MICROSATELLITES

PART A

4.1 Introduction

Blood groups have been used over the years to solve the problems of doubtful parentage and individual identification. Laboratory techniques involved in blood typing have improved considerably, in particular the use of microtitre plates and automatic dispensers for antisera have facilitated handling of large numbers of samples in routine blood typing laboratories.

DNA markers are much more efficient than blood typing in resolving identity because of their high level of polymorphism and ability to PCR (Jeffreys et al. 1991a). Microsatellites have been suggested to be the efficient tool for parentage determination, and their advantages over other genetic markers are explained in chapter 3. The conventional radioactive labelling and autoradiographic techniques used to characterize microsatellite loci, while more efficient than Southern blotting of RFLP loci, need to be substantially improved to facilitate routine parentage testing, which requires hundreds of thousands of marker genotypes.

In DNA typing laboratories, where many samples are to be analysed simultaneously, a robust and reliable method for typing of microsatellites has to be adopted for accurate and efficient genotyping. The narrow size range of alleles gives the potential for a number of loci to be analysed together by "multiplexing" (Chamberlain et al. 1990; Edwards et al. 1991; Morral and Estivill 1992). This is dependent on the choice of loci having similar reaction conditions with non interacting primers and the PCR products to be of different size for identification in the gel. Increasing the number of loci that can be analysed on each gel combined with automated typing would be a straight-forward way to improve the throughput.

4.1.1 Automated DNA typing

The number of loci analysed on a gel can be increased by co-electrophoresis in which the amplification of different loci is done in different PCR reactions and the products mixed together at a comparable concentration and loaded on a single lane for analysis. The use of multiplexing can also be extended to the co-amplification of several loci in a single PCR. This allows one to obtain more information from a single reaction, saving both time and reagents. The drawback of this method is that not all primer sets can be amplified together successfully, thus requiring extensive selection and testing to find compatible primer sets. Analysis of both multiplex PCR and multiplex electrophoresis is limited using a radioactive system as products cannot be distinguished if the loci overlap in size. Using the ABI automated DNA sequencer four distinct fluorescent dyes are used for analysis, thus co-electrophoresis and co-amplification are possible.

The advantages of analysing multiple loci by the use of an automated DNA sequencer was first described by Skolnick and Wallace (1988). Diehl et al. (1990) used a laser based automated DNA fragment analyser to size the PCR products from dinucleotide repeats amplified using fluorescently labelled PCR primers and demonstrated the feasibility of this approach in large scale genetic mapping studies. The Automated Laser Fluorescent (ALF) DNA sequencer (Pharmacia) uses a single fluorescent label (fluorescent isothiocyanate) to label the primers for use in the PCR reaction. Edwards et al. (1991) reported the use of three fluorescent dyes in typing trimeric and tetrameric tandem repeats in multiplex analysis. An internal size standard cocktail was prepared by amplifying specific alleles from individuals of known genotype. Improving upon the techniques of Edwards et al. (1991), Ziegle et al. (1992) reported the application of automated DNA sizing technology for genotyping microsatellites that use four colour fluorescent based techniques. PCR primers labelled with one of three fluorescent dyes having a different emission wavelength were included in the PCR, and the simultaneous analysis of microsatellite marker loci which overlapped in size within a single lane of a polyacrylamide gel was demonstrated. A fourth dye was used for a size standard in each gel lane, which is a Pst 1 digest of lambda phage.

The fluorescence emitted by each dye label differ in wavelength, and a specific filter is used to detect each dye label. Each of the two filter sets is composed of a fixed set of four filters, and a single filter set is used for an entire fragment analysis. Therefore, in an experiment the four dyes that are all detectable and distinguishable on a single filter set are to be used. Because there is an overlap in wavelength in the emission spectra of fluorescent dyes GENESCAN software uses a mathematical matrix to correct for the overlap before analyzing the data. So a matrix has to be set for each group of dyes and electrophoretic conditions.

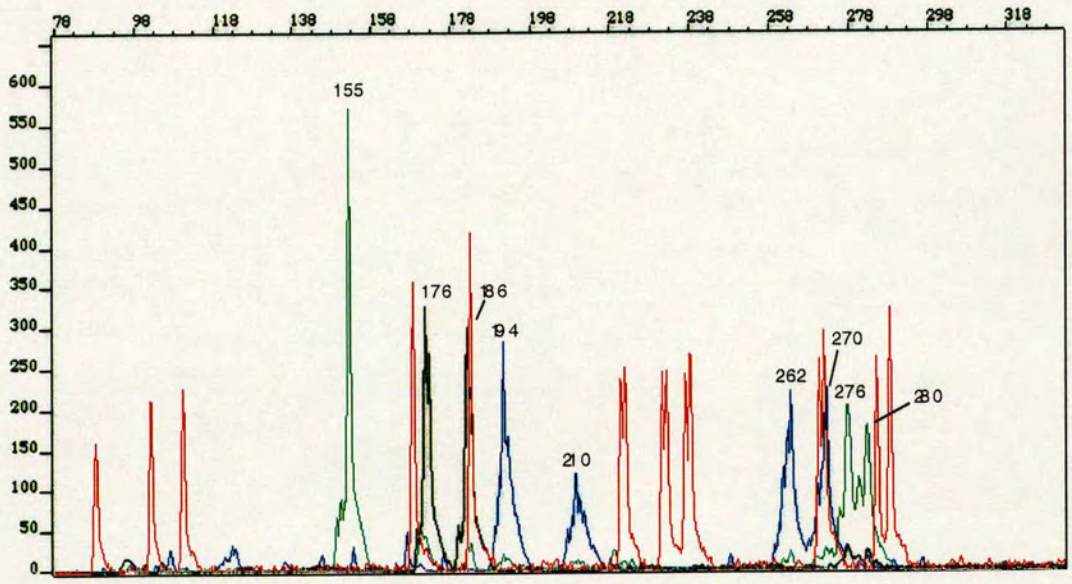
The Applied Biosystems model 373 DNA sequencer with computer software (GENESCAN 672) was developed to provide rapid and accurate sizing of alleles. The technique is based on the detection of fluorescence emitted from dye labelled DNA fragments during their electrophoretic separation in polyacrylamide gels. The fluorescence is excited by directing the beam of an argon-ion laser through a small area near the bottom of the separation gel. Emission from the dye-labelled DNA fragments passing through this area is focused by a collection lens through a four-wavelength-selectable filter and into a photomultiplier tube. Digitized signals from this detector are transfered directly into a computer for subsequent interpretation.

Recently, Schwengel et al. (1994) developed six sets of up to eight marker loci, the allele size ranges of which did not overlap within a set, and using a different dye for different sets, simultaneously analysed up to 24 microsatellite per gel lane and calculated the capacity of a single gel to be 864 genotypes. Reed et al. (1994) has suggested a potential daily throughput per ABI 373 DNA sequencer of 2,592 genotypes by running three gels per day.

4.1.2 Advantages of automated typing.

The major advantage of four colour fluorescent technology is the ability to use an internal lane size standard which minimizes the variation between lanes (Carrano et al. 1989). Fig 4.1 shows the data of precise allele size determination based on the migration pattern of the fluorescent labelled alleles using internal size standard in each lane. Another advantage is the ability to differentiate by colour between overlapping loci of similar size within the same region of the gel as

Figure 4.1 Electrophoretogram display of PCR products.



Electrophoretogram display of PCR products of five microsatellite loci analysed on a single gel lane. The markers are differentiated by the fluorescent labels. Red - internal size standard; Blue - FSHB and HEL6; Green - CYP21 and ETH131; Black - DRB3. The data are aligned by size (bp) and the size of the allele is digitized and is indicated at each peak. Overlapping products of different loci are easily identified.

indicated in fig 4.1. Fluorescent detection gives a measure of the peak height which is proportional to the signal intensity. This is of direct significance when typing samples of mixed origin in forensic science because the high sensitivity of the automated sequencer can quantitate even the slightest mixing of cells (Kimpton et al. 1993).

In the ABI 373 DNA sequencer the electrophoresis is continued until all distinguishable and detectable bands have passed the detector area. Fragments of interest are separated as they pass this point; once through it they are eluted of the end of the gel. Thus the whole length of the gel can be effectively utilized for multiplexing several loci. Also the internal standard in each gel lane eliminates the effects of variation in gel mobility.

The multiple co-amplification approach requires less template DNA per analysis than a regular single marker amplification. Fluorescently labelled PCR products and primers are stable indefinitely without deterioration if protected from light and stored at -20^{0}C. The advantages of using automated DNA sequencers are being extensively exploited by gene mapping laboratories and forensic laboratories, thereby allowing the comparison of results (Sullivan et al. 1992).

4.1.3 Limitations of autoradiography.

Conventional autoradiographic procedures for analysing amplified microsatellite loci have limitations for large scale genotyping, particularly since the major source of typing error is in the interpretation of autoradiographs (Beckmann et al. 1993). Analysing gels loaded with samples amplified at many marker loci requires several steps since manual interpretation of DNA patterns is complicated by the stutter bands, a characteristic of microsatellite marker loci (Smeets et al. 1989). With autoradiographic methods, multiple loading of samples is performed avoiding overlapping of fragments. Both the longest and shortest fragments of interest must be well separated and yet contained within the gel for analysis.

The standard approach to microsatellite analysis involves gel electrophoresis, which is complicated by problems with gel distortion (Lander 1991). Gel distortion can happen because of variation in mobility across the gel caused by temperature

gradients due to leakage of buffer or uneven current supply, or gel irregularities arising from preparation or handling. With this approach, allele identification and size calling can be problematic because of lane-to-lane variation of bands (Mayrand et al. 1992). In microsatellite analysis where alleles can differ in size by as little as one or two base pairs and also when stutter bands are generated (Gill et al. 1990, Evett and Gill 1991), it is necessary to evaluate the relative intensity of bands to distinguish homozygotes and heterozygotes. The non-linearity of the autoradiographic process makes it difficult to compare the individual band intensities. The use of radio-isotopes is hazardous when compared to the fluorescent dyes. Autoradiography also offers no easier way to deal with amplification efficiencies of PCR: faint bands can only be effectively visualized by re-exposing the gels to film or by re-running the gels with modified conditions. Throughput is limited as only one isotope is used. Rapid decay or degradation of labelled primers or PCR products limits using them for further analysis.

The aim of the work described in this chapter is to explore a multiplex automated system using microsatellites for routine typing in parentage testing laboratories.

4.2 MATERIALS AND METHODS

Five microsatellite loci were amplified from 36 DNA samples: three of the loci had a size of approximately 150-200bp and the other two were in the 260-275bp range. The primers for the loci which overlapped in size were labelled with different fluorescent dyes, thus facilitating analysis on a single lane on the gel. The dyes used were FAM and JOE for HEL6 and CYP21 (260-275bp) respectively and FAM, TAMRA and HEX for FSHB, DRB3 and ETH131 (150-200bp) respectively. Table 3.3 indicates the fluorescent labels and the product size for each marker. PCR conditions were as described in Table 3.3.

The annealing temperature selected for each marker was such that it eliminated non-specific annealing or artifact bands. The reaction conditions were similar for three of the markers, CYP21, FSHB and DRB3, so these could be potentially amplified in one reaction, while ETH131 & HEL6 required different

conditions and thus had to be amplified separately. PCR amplification was carried out as described before, with 0.5uM of primer for all the markers except for FSHB for which 1uM of primer was used. PCR products from the triplex reaction and HEL6 were diluted four times with water and the product of ETH131 six times. Before loading to the gel, 1ul of the diluted PCR product from each of the three reactions was mixed together with 0.5ul of the internal lane size standard and 4ul of deionised formamide.

The preparation of the gel and electrophoresis methods are included under the general materials and methods, section 2.7.3.

4.3 RESULTS

Both co-electrophoresis and co-amplification were developed and evaluated with the five markers. Initially radioactively labelled PCR products of microsatellite loci CYP21 and DRB3 were co-electrophoresed on acrylamide gels. Good resolution for both products with clear discrimination between loci was obtained. Also co-electrophoresis of the PCR products of CYP21 with those of FSHB and ETH131 also gave good results. Similarly co-electrophoresis of the PCR products of HEL6 was tried with those of all the three markers of smaller product size - DRB3, FSHB and ETH131 separately with good resolution of products.

Then co-amplification of radioactively labelled PCR products of different markers was attempted. Even though DRB3 could be amplified with 1mM $MgCl_2$, co-amplification was tried with CYP21 by increasing the magnesium concentration to 1.25mM. For PCR 0.5uM of each primer was used and annealing was done at 60^0C. Clean products with good resolution were seen for CYP21 and DRB3. A similar approach was made with CYP21 and FSHB. Good amplification was seen for CYP21 but little product was found for FSHB in the co-amplification. When the concentration of the primers for FSHB was increased to 1uM, however, reasonable amounts of product were obtained. Co-amplification of HEL6 and ETH131 was not possible because the annealing temperatures (60^0C and 55^0C) and magnesium concentrations (1.5mM and 1.25mM) were different for these loci. The co-amplification and co-electrophoresis of more than two markers were not possible

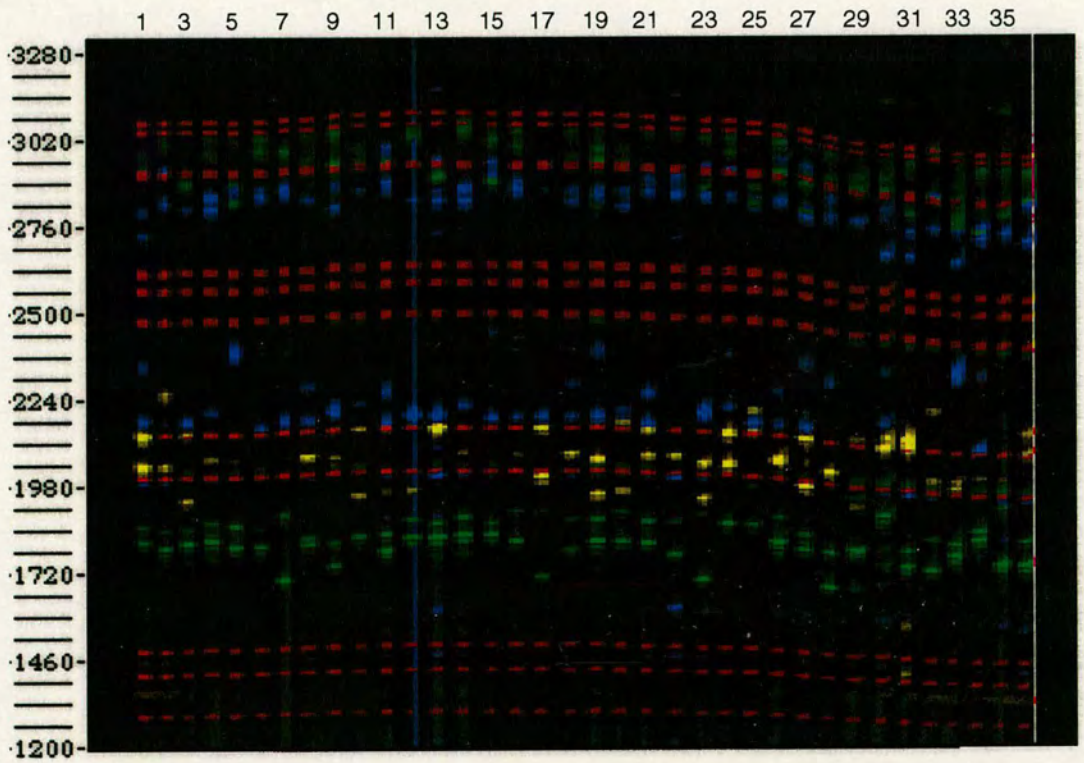
with radioactively labelled PCR because the products overlapped in size and could not be distinguished.

As a next stage, independent PCR amplification of the five markers, CYP21, DRB3, FSHB, ETH131 and HEL6 were tried with fluorescent tagged primers and were co-electrophoresed with products from the separate reactions on an Applied Biosystems 373 DNA sequencer. Even though the products for DRB3, FSHB and ETH131 overlapped in size, the products could easily be identified through the different colours (yellow, blue and green) of fluorescent dyes attached to the primers. Similarly, the products of CYP21 and HEL6 were distinguished by using different fluorescent tags.

Co-amplification of CYP21 and DRB3 was tried and co-electrophoresed with the products of FSHB, HEL6 and ETH131. CYP21 and DRB3 gave good PCR products on co-amplification, and on electrophoresis with other markers good resolution was obtained on the gel for all markers. Because the PCR conditions for amplification of FSHB were similar to those of CYP21 and DRB3, co-amplification of these three loci were tried in a single PCR tube with a primer concentration of 0.5uM for each locus. Little product was obtained for FSHB. However, as with the use of radioactivity when the primer concentration of FSHB was increased to 1uM in co-amplification, similar amounts of products were obtained for all the loci. Thus the three markers CYP21, DRB3 and FSHB could be co-amplified successfully. Because of the difference in annealing temperature and magnesium concentration for the markers HEL6 and ETH131, co-amplification of these markers was not successful with any of the markers included in the triplex reaction mentioned above.

Thus with the five markers included in this study, multiplexing by co-amplification of the markers CYP21, DRB3 and FSHB and then co-electrophoresing with the products of HEL6 and ETH131 loci which were amplified separately enabled analysis of the five markers on a single lane in the gel. Fig 4.2 shows the multiplexed gel with PCR products from 36 samples amplified for the five microsatellite loci, DRB3, CYP21, FSHB, ETH131 and HEL6. The PCR products for CYP21 (green) and HEL6 (blue) were within the size range of 260-275bp but could be distinguished through the use of different dyes. Similarly DRB3 (yellow),

Figure 4.2 View of multiplexed gel with five mircosatellite markers



Five microsatellite markers are analysed on 36 different DNA samples. The five loci are amplified in three separate PCRs. Three fluorescent labelled PCR primer sets FAM (blue), HEX and JOE (green) or TAMRA (yellow) each specific for a different marker are used.

ETH131 (green) and FSHB (blue) were differentiated by the dyes even though they overlap in size. The fourth dye, ROX (red) was used for an internal marker. Thus the use of four dyes permitted precise allele determination for DNA fragments.

Fig 4.1 is an "electrophoretogram" of a single gel lane loaded with PCR products for the five loci. Each peak in the electrophoretogram shows the relative intensity of the DNA fragments, and faint bands could be better visualised by adjusting the fluorescence scale. Alleles were defined as the highest peaks within the expected allele size range. Some interference was observed between the yellow and green signals from the PCR products of ETH131 and DRB3 loci when overlapped in size because the concentration of PCR products of ETH131 was very high when compared with DRB3.

The analysis performed using fluorescently labelled primers eliminated the need for radioactively labelled materials. In addition because of the extreme sensitivity of the the ABI automated sequencer, only one fifth of a microlitre of fluorescently labelled PCR products was used when compared to 5ul of radioactive products allowing low cycle number for PCR amplification and thus reducing the possibility of artifacts.

Therefore multiplexing, with co-amplification and co-electrophoresis performed separately or together could be used to substantially reduce the cost of reagents and labour in routine parentage testing laboratories.

4.4 DISCUSSION

A high throughput of genotyping with microsatellite loci has been limited by the technology used to analyze the PCR amplified products, allele scoring accuracy and efficiency. Use of fluorescently labelled PCR products and the automated fragment estimation and sizing using the DNA sequencer offer significant improvements over radioactive labelling methods to overcome these limitations (Ziegle et al. 1992). The multiplex DNA amplification procedure provides a rapid and simple method to analyze simultaneously a large number of samples with many loci. In this chapter a multiplex amplification system in conjunction with automated fluorescence based detection was developed with the highly polymorphic

microsatellite loci CYP21, DRB3, FSHB, ETH131 and HEL6 which could be used in routine parentage verification in cattle.

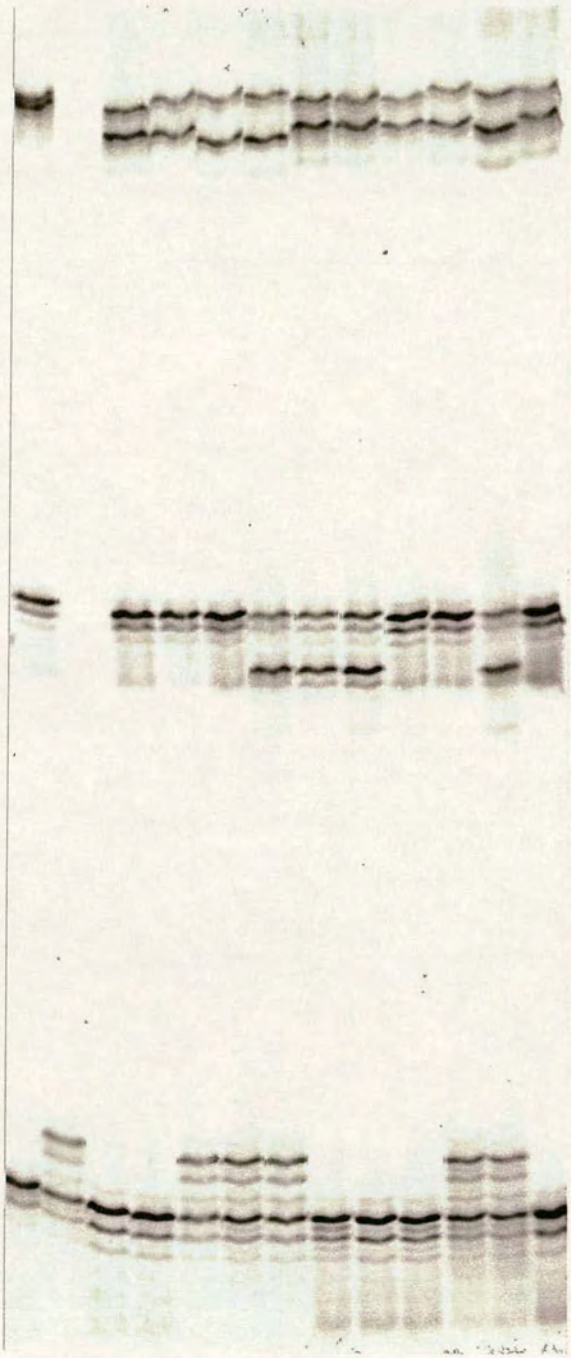
With radioactive labelling, the products of only two of the markers could be co-electrophoresed because of the overlap in size. The amplification conditions of the five markers using radioactivity showed that three of the markers could be co-amplified under standard PCR conditions. Efficient amplification of the three markers in the automated sequencer, CYP21, FSHB and DRB3 was achieved by adjusting individual primer and magnesium concentration. The narrow size range of the markers used, together with the different fluorescent tags on the loci with overlapping size ranges, facilitated the discrimination of the products. However, the PCR products of HEL6 and ETH131 were co-electrophoresed with the products of the triplex reaction allowing simultaneous analysis of the five markers in a single gel lane. The PCR products were diluted to different concentrations to give bands of equal intensity.

Published reports indicate that PCR products of many bovine microsatellite loci range in size from 150-200bp. The advantage in the use of primers labelled with different fluorescent dyes is that the problem of overlaping in the size of these PCR products can be overcome and there accurate sizes determined.

The advantage of co-electrophoresis is that optimum reaction conditions for all the markers can be used, so that the maximum number of samples can be amplified. But in co-amplification, the reaction conditions for each marker are carefully adjusted to satisfy amplification requirements. Reaction conditions should be optimal for each marker to avoid inadequate amplification of the samples.

The number of radiolabelled products analysed on a single gel can be considerably improved by multiple loading. In multiple loading, successive samples are loaded on to the gel at intervals so that the samples of each loading are well separated from the other and are easily readable. For example, in an electrophoresis system with 48 gel lanes, with three successive loadings 144 samples can be genotyped. Fig 4.3 indicates a gel with three multiple loadings. Multiple loadings can be done either with samples amplified by the same loci so that all the products are of the same size or amplified across different loci for the products to be of

Figure 4.3 Autoradiograph of the gel showing co-electrophoresis of PCR products



PCR products of the marker IDVGA90 labelled with ^{33}PATP are loaded at successive intervals allowing clear separation between products after each loading.

different size. Thus the whole of the gel can be effectively used for genotyping a large number of samples. However, a disadvantage in this method is that an accurate estimate of the sizes of the PCR products from successive loadings is not possible because the size standard loaded at one time is not suitable for other loadings. This system is only useful for comparison of inheritance of alleles within families.

Because multiplex PCR products are synthesised under standardised amplification conditions, it may be possible to relate the fluorescent intensity of specific alleles at different loci (Edwards et al. 1991). Fluorescence signals are linear over a much greater range of intensities than conventional autoradiography (Ziegle et al. 1992). Therefore, fluorescence based detection systems are more useful than autoradiography for direct quantitation of PCR products. Also these systems can detect efficiently a greater range of intensity of signal. On the other hand, in autoradiography weak signals are often lost. For gels loaded with radioactive PCR products of markers FSHB and ETH131, where multiple "stutter" bands are found, manual interpretation of patterns was difficult following autoradiography. However fluorescent DNA typing methods distinguish between alleles and PCR artifacts by revealing the relative intensity of the bands and thus overcome most of the difficulties in typing microsatellite markers. This is because major peaks are obviously higher than the "stutter" bands.

With automated DNA sequencers, even though, the mobility of the PCR products was affected by electric field or gel composition unhomogeneity, uneven heat distribution (smiling) and high sample ionic strength (salt effect) the sizing of the PCR products is not affected because it is compensated for by the internal size standard. Such factors may cause two identical length PCR products in adjacent lanes to migrate differently or cause otherwise different length PCR products to migrate similarly in radioactive gels. This problem is obviated with the use of an internal lane standard because both the sample and the standard experience the same electrophoretic environment and thus migrate relative to one another.

In the present study the fluorescent dyes HEX and JOE are green dyes but use different filter sets to detect the dye label. Because most of the dyes, FAM,

HEX, TAMRA and ROX were detected by filter set A, the gel was analysed using this filter set. So all the dyes except JOE are clearly seen in the gel image.

For microsatellite markers like FSHB, where some alleles differ by only 1bp, precise allele designation was possible with automated typing which was not possible by autoradiography. This accurate detection of alleles was consistent with the size estimation accuracy reported by Carrano et al. (1989). Schwengel et al. (1994) concluded fluorescent based semiautomated genotyping was at least as accurate as the standard radiolabelling technique. In the multiplex system developed here the errors found in sizing of the products were less than 1bp with the use of the multicolour fluorescence based technique, suggesting that this methodology scores microsatellite alleles in a highly reproducible manner. The system is much more efficient than the conventional radiolabelling technique for which the repeat typing showed an error of +/-2bp. Schwengel et al. (1994) investigated the variation within and between gels using fluorescence based technology, and observed a maximum size variation of 0.47bp over 571 comparisons.

In this study, the sizing of fluorescent PCR products of grossly different concentrations was potentially complicated by sample overloading, which caused spectral interference between the dye labels during analysis. There was some interference in the detection of the overlapping products (DRB3, FSHB and ETH131) because the concentration of PCR products of ETH131 was comparatively high when compared to that of DRB3 and FSHB at equal dilutions. The peaks in the green and blue electrophoretogram were seen in the yellow, but Schwengel et al. (1994) observed no interference between overlapping products using the four dyes. It is impractical to estimate the concentration of each of the fluorescent products so as to adjust the concentration of individual samples to be pooled during large scale typing procedures. This was accomplished in the present study by viewing the electrophoretogram for band intensities initially so as to assess the concentration of products for each marker.

Development of multiplex system with different microsatellite loci for parentage testing would increase the number of samples genotyped per day considerably. The commercial "StockMarks" kit available, from Perkin Elmer has

11 microsatellite markers which are co-amplified in three multiplex sets and electrophoresed in a single lane. In this kit, not all the markers amplify very well. Moreover, for the markers which can be amplified, the peak intensities for different products were so different that the scale in the electrophoretogram for one marker could not be used for another. This problem was often complicated by the appearance of peaks from the blue and green dyes into the yellow. Allele frequencies provided for the markers in the "StockMarks" kit were based on data from the Holstein Friesian breed of cattle. Some markers were found monomorphic when tested on other breeds and some markers less polymorphic, so when used in paternity testing, the probability of exclusion calculated was low when compared to the panel of markers used in this thesis.

PCR reactions done in 96 well microtitre plates instead of in PCR tubes speeds up the process of setting up the reactions. Use of robotic equipment for aliquoting reagents in microtitre plates would further improve the technique. If these methods are applied in a parentage testing laboratory, tremendous saving of time and effort would be achieved by a one day multiplex amplification protocol. Thus, the combination of multiplex microsatellite PCR and automated fluorescent detection results in a rapid and powerful DNA profiling technique with considerable potential for high throughput.

PART B

4.5 Practical applications using microsatellites for identity and paternity testing

Blood typing for parentage testing has been offered as a service to breeders and herdbook societies by laboratories throughout the world since about 1950. Parentage testing is carried out to verify the relationship of animals for pedigree registration, to determine the father in case of multiple inseminations, and to confirm ancestry in animals originating from embryo transfer. When a parentage error is detected the breeder then has to identify the correct parents before the calf can be registered, and finding the source of error can be difficult in some cases especially when artificial insemination is involved. Using the traditional blood group and enzyme polymorphisms a fresh sample of blood is required, thus necessitating access to the individuals involved. However, with the introduction of DNA methods it is now possible to demonstrate the origin of semen and tissue samples. This approach also enables the identity of semen to be checked.

Two cases of individual identification and parentage determination using five microsatellite loci used in this work and "StockMarks" kit are described below. The microsatellite polymorphisms were detected either by autoradiography or by automated fluorescence detection on the Applied Biosystems 373 DNA sequencer.

Case 1

4.5.1 Determination of parentage of four calves.

Four crossbred (Hereford X Friesian) in calf heifers belonging to a farmer were reported stolen from his farm. Six months later, four cows with four calves were found in a field and it was suspected that these might be the stolen animals. The lost heifers had been mated with a Simmental bull owned by the farmer. Blood samples were collected from the bull, the four cows and the four calves to test whether these calves were out of the four cows by the Simmental bull. Relationship between the cows, calves and the bull were to be established by both blood typing and microsatellite analysis with five markers.

4.5.1.1 Methods

DNA was extracted from blood samples of the bull, four heifers and the four

calves as described in Chapter 2. The samples were blood typed and analysed for transferrin polymorphism. Five microsatellite markers, CYP21, DRB3, ETH131, FSHB and HEL6, were radioactively labelled and analysed by autoradiography. Details of the markers, primers and method of analysis is given in Table 3.3.

4.5.1.2 Results

The allele sizes for the five microsatellite markers were determined manually by comparing with the size standard. Autoradiographs representing the nine samples for the markers CYP21 and DRB3 are shown in Fig 4.4. All the markers were polymorphic. Genotypes of the calves 1, 2, 3 & 4 and of the cows A, B, C & D are shown in Table 4.1.

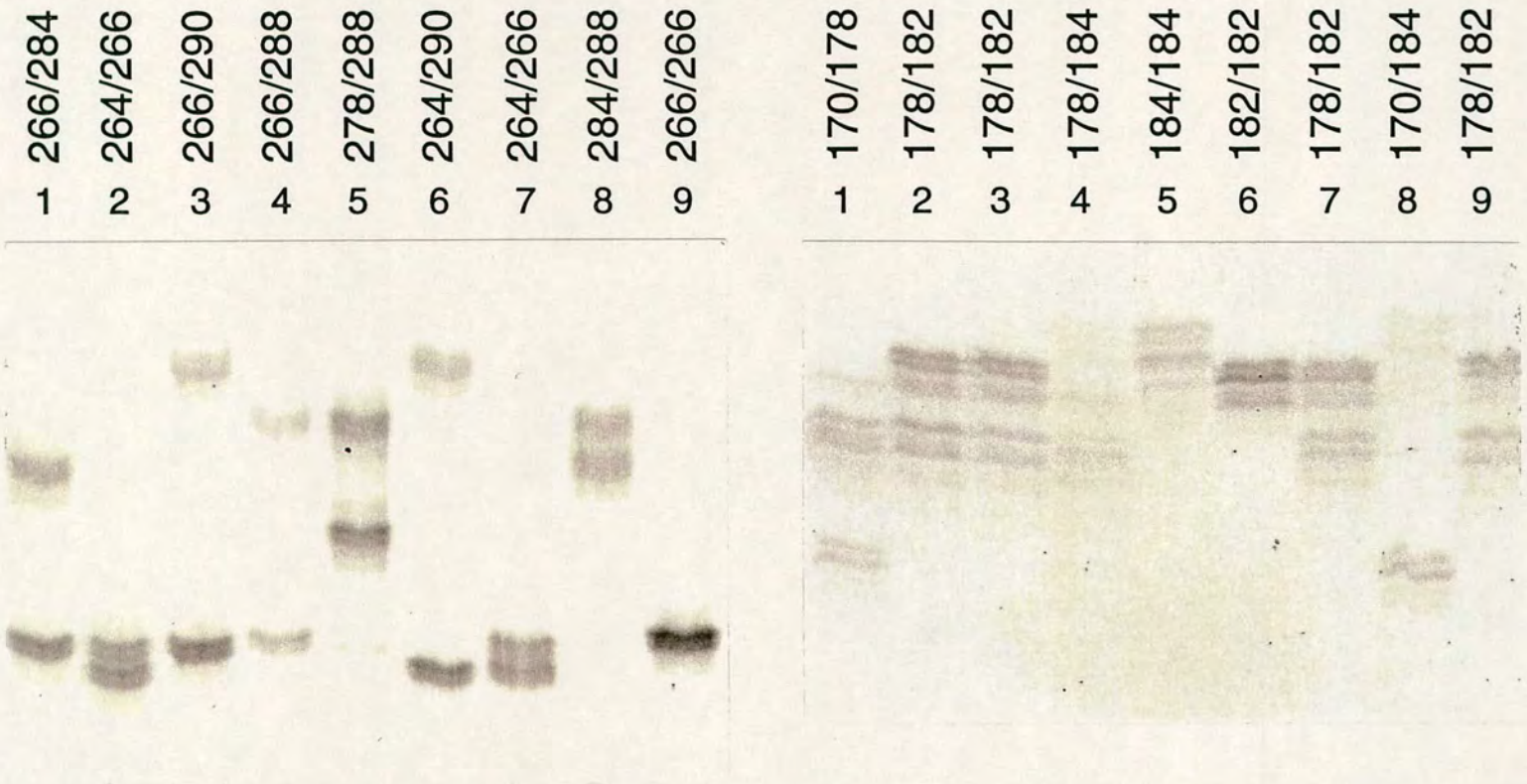
It was found that the calves had received alleles from one or the other of the dams and from the bull. Following the inheritance of alleles for the five markers, it is suggestive that the four calves could be sired by the same bull and the four cows in question A, B, C & D could be the dams of the calves 4, 3, 2 & 1 respectively. Results of blood typing and transferrin polymorphism analysis are shown in Table 4.2. These results also agree with that from the microsatellite markers.

Using the genotypes of the bull and the cows, the calves were matched to the probable dams. Frequency of alleles for different microsatellite markers was based on the data from the 20 unrelated animals of these breeds (see Table 3.5 - 3.9) depending on whether the allele is inherited from the Simmental bull or Hereford X Friesian cows. The highest frequency of the alleles among Herefords and Friesians was used for calculation.

The likelihood of finding the genotypes of these calves, the cows and the bull at random in the population on the basis of microsatellite markers is given in Table 4.1. Taken together, all the five markers the probability of the genotypes of the four calves to occur at random in the population are 5.27×10^{-14}, 3.08×10^{-12}, 5.62×10^{-15}, and 8.15×10^{-13} respectively.

Given the genotype of the calf, the probability of another dam and sire in the population having a calf with the same genotype is also calculated based on the results of microsatellite markers and are shown in Table 4.1. These probabilities based on the five markers for the calves 1, 2, 3 & 4 are 1.67×10^{-12}, 5.58×10^{-11},

Figure 4.4 Autoradiograph of the gel with PCR products confirming parentage



Autoradiograph of the gel showing PCR products amplified with markers CYP21 and DRB3 on DNA samples from four calves, four dams and a bull to confirm parentage.Lanes 1,2,3,4 are of the calves, 5,6,7,8 of the dams and 9 of the bull.

Table 4.1 Probabilities of matching of genotypes of the calves, dams and the sire in the population

	CALVES				DAMS				SIRE	
Microsatellite marker	1	2	3	4	A	B	C	D	**c)**	**d)**
CYP21	266/284	264/266	266/290	266/288	278/288	264/290	264/266	284/288	266/266	
Probability a)	0.0125	0.025	0.001	0.0075	0.011	0.0025	0.0437	0.0093	0.010	0.19
b)	0.0250	0.050	0.002	0.0150						
DRB3	170/178	178/182	178/182	178/184	184/184	182/182	178/182	170/184	178/182	
Probability a)	0.0006	0.0075	0.0012	0.0031	0.0156	0.0025	0.0075	0.0031	0.0002	0.049
b)	0.0012	0.015	0.0025	0.0062						
ETH131	148/158	148/154	154/162	154/162	143/162	154/162	148/158	148/158	148/154	
Probability a)	0.0037	0.0031	0.0003	0.0003	0.0001	0.0005	0.0031	0.0031	0.0038	0.01
b)	0.0075	0.0062	0.0005	0.0005						
FSHB	196/210	198/210	202/212	198/212	192/198	190/202	198/200	196/210	210/212	
Probability a)	0.0019	0.0106	0.015	0.0425	0.0210	0.0225	0.0740	0.0037	0.0025	0.005
b)	0.0037	0.0212	0.030	0.0850						
HEL6	259/261	257/265	257/257	261/271	263/271	257/257	257/265	259/265	257/261	
Probability a)	0.001	0.0005	0.001	0.0055	0.00275	0.010	0.005	0.005	0.0001	0.0002
b)	0.002	0.001	0.001	0.0027						
Overall a)	5.27×10^{-14}	3.08×10^{-12}	5.62×10^{-15}	8.15×10^{-13}						9.31×10^{-11}
b)	1.67×10^{-12}	5.58×10^{-11}	7.5×10^{-14}	2.17×10^{-11}						

a) Probability of genotypes occurring at random, b) Probability of finding another sire and dam which could give the genotype found in calf, c) Probability of observed sire genotype occurring at random, d) Probability of finding another sire which could contribute the non-maternal alleles in the calves.

Table: 4.2 - Results of blood typing in confirming parentage of four calves

		B system	C system	Tf	pAlb
V19832	Bull K70 W7781	$BI_1/O_1I'_1$	C_1X_2	AD_2	
V19824	Calf 1 (Red & White)	$BI_1/O_1A'1$	C_1X_2	D_2D_2	
V19831	Dam No. 7/GJ630	$GY_2E'2Q'/O_1A'_1$	$C_1R_1W_2$	AD_2	SS
V19825	Calf II (Black & White)	BI_1/BGO_2	$C_1R_1W_2X_2$	AA	
V19830	Dam No. 6/GJ630	$BGKO_1Y_1A'_2G'_2O'_2P'/BGO_1$	$C_1R_1W_2X_2$	AA	SS
V19826	Calf III (Black & White)	$BI_1/O_2D'F'_2G'_2O'_2$	C_1WX_2	D_1D_2	
V19829	Dam No. 4/GJ630	$O_2D'F'_2G'_2O'_2/Y_1D'I'$	C_1W_2	AD_1	SS
V19827	Calf IV (Black & White)	$BI_1/BO_2Y_2G'_2P'_2Q'$	C_1	AD_2	
V19828	Dam No. 5/GJ630	$BO_1Y_1(A')G'_2P'_2Q'/BO_1Y_1G'P'Q'/Q'$	C_1W	D_1D_2	SS

7.5 X 10^{-14} and 2.17 X 10^{-11} assuming that all the alleles are correctly identified.

Assuming that all the four calves are matched to their probable dams, the probability of finding another sire to have four calves with the given genotypes is calculated for each marker separately and are 0.19, 0.049, 0.01, 0.005 and 0.0002 respectively for CYP21, DRB3, ETH131, FSHB and HEL6.

These probabilities indicate that the chance of finding another sire/dam combination or another sire in the population to contribute the non-maternal allele for the four calves would be much more than 1 in 10^{12}.

4.5.1.3 Discussion.

It is not surprising that the probabilities calculated for the five microsatellite markers are very high because of high levels of polymorphism. The frequency of microsatellite alleles are not exact, because the frequencies used for calculation were taken either from Hereford or Friesian (which ever is highest) and the Simmental depending on the alleles inherited. But even by considering the most frequent allele, the value calculated is very high. Accounting for the errors in genotyping would further increase the probability.

With over 700 microsatellite markers available in cattle (Barendse et al. 1994, Bishop et al. 1994), a panel of markers which will be polymorphic and informative for each breed can be selected. The allele frequency for microsatellite markers being highly variable between breeds, it is necessary to use the allele frequencies in each breed for calculation of probabilities. This case is typical in having a sire of one breed (beef breed) and the dam from a cross between another two breeds. Considering the worst situation, the highest value of allele frequency between breeds is taken for calculation.

Case 2

4.5.2 Verifying the identity of bovine semen using DNA microsatellite markers.

A Charolais heifer was presented to the herd society for registration. The reported identity of its sire and dam were available, but the results of blood typing confirmed the dam but not of the sire. The breeder involved was certain about the sire because he had kept the actual semen straw used for insemination. So the

identity of the semen supplied in the straw was to be confirmed on the basis of the semen samples obtained from this bull. Although the breeder had the actual semen straw used, it had been kept in dubious conditions and therefore was not a source of DNA. However, it was possible to trace several straws from the same batch collected and filled on the same day as the straw which was disputed, in addition the bull in question had been owned by the Irish Department of Agriculture who were able to provide semen from several collections over several years from the bull. The semen sample provided was checked to see whether the suspected straws contained semen from the same individual as the undisputed straws using microsatellite markers.

4.5.2.1 Methods.

Samples comprised a straw from the disputed batch, 11 different undisputed semen collections for the same bull and semen from 17 other Charolais bulls used by the North Western Cattle Breeding Society (NWCBS). In addition DNA was available for a further 20 unrelated Charolais sires used in the UK.

DNA was extracted from blood and semen samples as described in Chapter 2. Two approaches to microsatellite DNA analysis were used, either by labelling PCR primers by ^{33}P or by fluorescent labelling. The 28 DNA samples, 10 from undisputed semen straws of the bull, 17 straws from unrelated Charolais bulls and 1 straw from the same collection as the disputed straw were genotyped for 16 DNA microsatellite markers. Five microsatellite DNA markers, CYP21, DRB3, ETH131, FSHB and HEL6 were radioactively labelled and analysed by autoradiography, the details of the markers, primers and method of analysis are given in Table 3.3. A further 11 microsatellite markers were from the ABI "Stockmarks" Kit (TGLA48, TGLA263, TGLA53, MGTG7, TGLA57, TGLA73, MGTG4b, AGLA293, TGLA227, TGLA126, and TGLA122). These markers were co-amplified in three multiplex PCR sets, two each with four markers and one with three markers and results were analysed on an ABI model 373 automated DNA fragment analyser. The details of the markers are as follows.

Marker	Dye	Colour	Allele size Range
TGLA48	TET	Green	69-87
TGLA263	TET	Green	100-136
TGLA53	TET	Green	144-190
MGTG7	TET	Green	273-329
TGLA57	FAM	Blue	80-104
TGLA73	FAM	Blue	110-128
MGTG4B	FAM	Blue	134-164
AGLA293	FAM	Blue	196-260
TGLA227	HEX	Yellow	78-104
TGLA126	HEX	Yellow	109-127
TGLA122	HEX	Yellow	130-164

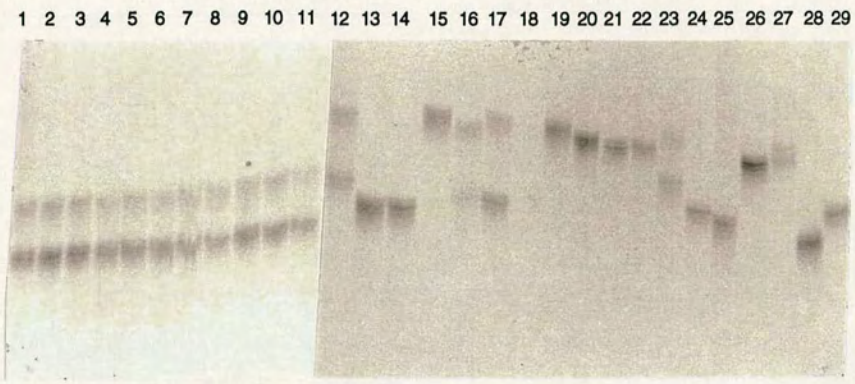
4.5.2.2 Results

The allele sizes for the DNA markers analysed using radioisotope were determined manually by comparison with a sizing standard, and an example result is shown in fig 4.5. The fluorescently labelled markers were analysed automatically and the gel image is as shown in fig 4.6. All the microsatellite markers showed high levels of polymorphism in the 18 Charolais sires used by the NWCBS. The alleles found and frequencies for markers analysed by autoradiography are shown in Table 4.3a, whereas allele frequencies found in 20 unrelated Charolais animals across the UK are shown in Table 4.3b.

Only the 18 NWCBS Charolais sires were genotyped for the 11 "StockMarks" markers, which were analysed using the ABI 373 DNA sequencer. Allele frequencies for these are shown in Table 4.3c. Repeat typing showed errors of +/- 2bp using the radioisotope method, while replicates using the automated procedure had an error of less than 1bp.

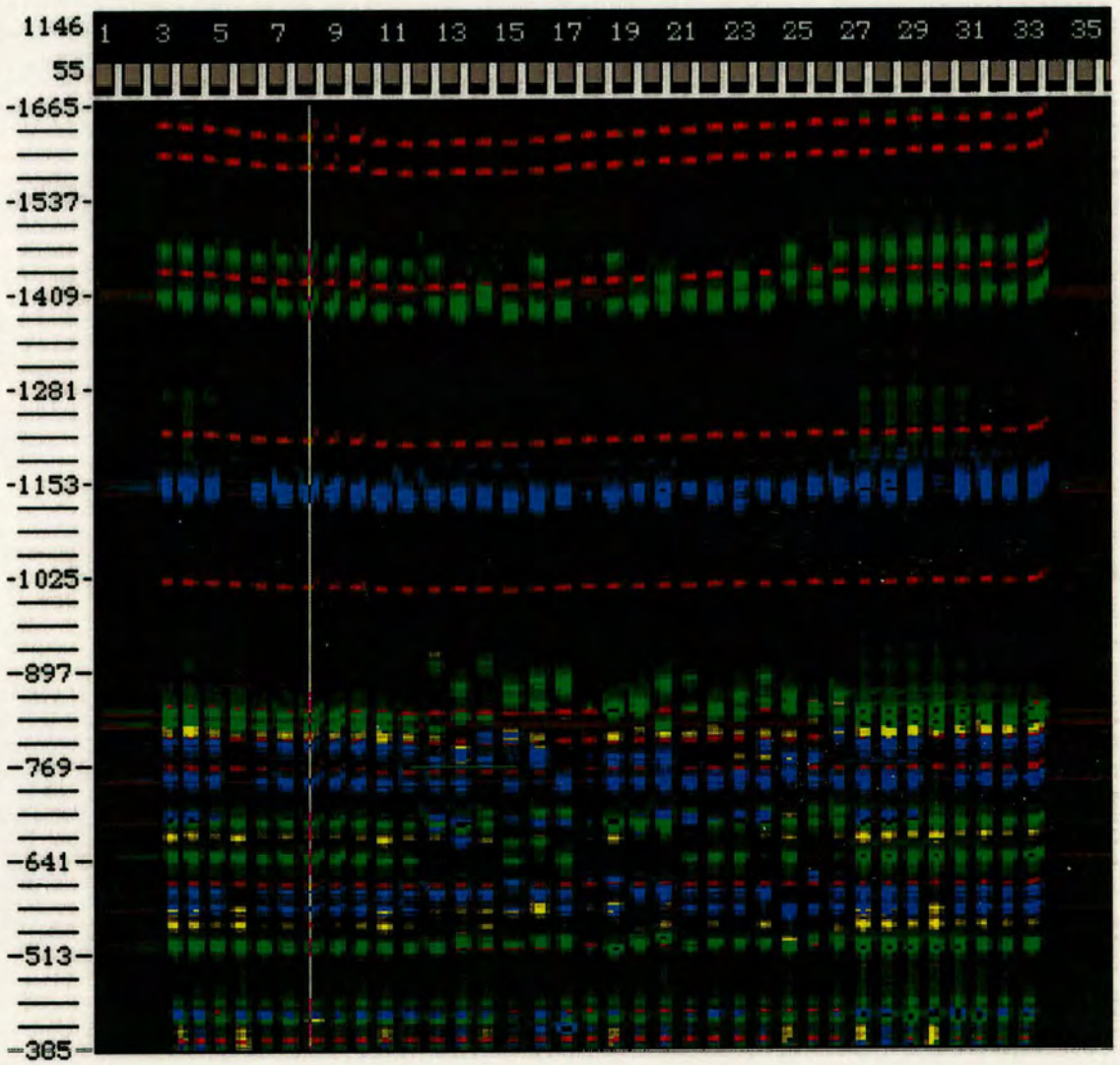
Using the genotype for the disputed bull matching probabilities were

Figure 4.5 Autoradiograph of PCR products proving identity of a disputed semen



Autoradiograph of PCR products of Charolais semen samples amplified by DRB3. Lane 1 disputed semen sample, lanes 2-11 other samples of the same batch of semen, lanes 12-29 several DNA samples from unrelated animals .

Figure 4.6 Multiplexed gel with microsatellite markers in "StockMarks" kit



Eleven markers in the "StockMarks" kit are analysed on 18 DNA samples. First lane is the DNA of the disputed semen sample and the next 10 lanes are samples from 10 straws of the same batch and the last 8 lanes are the DNA of the semen from 8 collections from the same bull. Lanes 12-28 are DNA samples from unrelated animals.

TABLE 4.3 Alleles and Frequencies for Microsatellite markers in Charolais samples

A) NWCBS animals

HEL 6		CYP21		DRB3		FSHB		ETH131	
250	<0.01	259	0.176	151	<0.01	168	0.031	125	0.031
252	0.083	261	<0.01	153	0.205	170	<0.01	127	<0.01
254	0.333	263	0.088	155	0.147	172	0.375	129	0.093
256	0.277	265	0.088	157	0.058	174	<0.01	131	0.281
258	0.027	267	0.411	159	<0.01	176	<0.01	133	0.093
260	0.055	269	<0.01	161	0.029	178	<0.01	135	<0.01
262	0.111	271	0.205	163	0.029	180	0.062	137	<0.01
264	0.055	273	<0.01	165	<0.01	182	0.343	139	0.062
266	0.027	275	0.029	167	0.297	184	<0.01	141	0.125
268	<0.01			169	0.235	186	0.062	143	0.218
270	0.027					188	0.093	145	<0.01
						196	0.031	147	0.031
								149	0.062

B) 20 unrelated animals

252	<0.01	259	0.050	151	0.200	186	<0.01	147	0.050
254	<0.01	261	<0.01	153	0.050	188	0.050	149	0.275
256	0.10	263	0.050	155	<0.01	190	0.375	151	0.125
258	0.125	265	0.075	157	<0.01	192	0.025	153	0.025
260	0.125	267	0.225	159	<0.01	194	<0.01	155	0.050
262	0.40	269	0.050	161	<0.01	196	<0.01	157	0.075
264	<0.01	271	0.350	163	<0.01	198	<0.01	159	0.150
266	<0.01	273	0.125	165	<0.01	200	0.050	161	0.150
268	0.025	275	<0.01	167	<0.01	202	0.300	163	0.025
270	<0.01	277	<0.01	169	<0.01	204	<0.01	165	0.075
272	0.05	279	0.025	171	0.025	206	<0.01		
274	0.15	281	0.025	173	0.175	208	0.200		
276	0.025	283	0. 025	175	0.100				
		285	<0.01	177	0.050				
				179	0.075				
				181	0.075				
				183	0.050				
				185	0.100				
				187	<0.01				
				189	0.10				

C) NWCBS animals

TGLA 227		TGLA 126		TGLA 122		TGLA 57		TGLA73	
75	0.070	115	0.692	142	0.176	76	0.028	115	0.029
77	0.030	117	<0.01	144	<0.01	78	0.028	117	0.029
79	<0.01	119	<0.01	146	0.058	80	<0.01	119	0.029
81	0.070	121	0.231	148	<0.01	82	<0.01	121	0.617
83	0.200	123	0.077	150	0.264	84	<0.01	123	0.294
85	<0.01			152	0.264	86	<0.01	125	<0.01
87	<0.01			154	0.118	88	0.028	127	<0.01
89	0.367			156	<0.01	90	0.361		
91	0.200			158	0.029	92	<0.01		
93	<0.01			160	0.088	94	<0.01		
95	0.030			162	<0.01	96	<0.01		
97	<0.01					98	0.444		
99	0.030					100	0.083		
						102	<0.01		
						104	0.028		

MGTG 413		AGLA 293		TGLA 48		TGLA263		TGLA53		MGTG7	
131	<0.01	221	0.029	72	<0.01	109	0.205	152	<0.01	290	<0.01
133	0.500	223	<0.01	74	0.219	111	<0.01	154	0.059	292	0.529
135	<0.01	225	<0.01	76	0.585	113	<0.01	156	<0.01	294	<0.01
137	0.055	227	0.029	78	0.195	115	0.058	158	0.176	296	0.118
139	<0.01	229	<0.01	80	<0.01	117	<0.01	160	0.206	298	0.059
141	<0.01	231	0.559			119	<0.01	162	0.118	300	0.088
143	0.222	233	0.382			121	0.529	164	0.118	302	0.029
145	0.083					123	<0.01	166	<0.01	304	<0.01
147	<0.01					125	0.205	168	0.147	306	<0.01
149	0.055					127	<0.01	170	0.118	308	0.147
151	<0.01					129	<0.01	172	<0.01	310	0.029
153	0.083					131	<0.01	174	<0.01	312	<0.01
155	<0.01							176	0.059		
								178	<0.01		

calculated for the 5 markers genotyped using autoradiography for the 17 NWCBS Charolais sires and also for 20 other unrelated animals (Table 4.4a). For the NWCBS data the overall matching probability was $7.6x10^{-8}$. Using the autoradiographic approach it is not always possible to make a precise identification of an allele, but the error will not be greater than mistaking the true allele for the next bigger or smaller. Recalculating the probability by pooling the frequency of the observed allele size with that of the next largest and the next smallest takes into account the error in determining the alleles involved in the test sire and also errors in determining the alleles in the sample panel, in fact this calculation will overestimate the error and hence the matching probability. Using this approach the matching probability drops to $6.86x10^{-5}$. Calculating the matching probability using the allele frequencies for the unrelated animals from other parts of the UK, the probability of finding an animal with the same genotype is less ($1.57x10^{-15}$) assuming all alleles are correctly identified, or $4.1x10^{-9}$ after accounting for errors.

The probability of finding another animal with the genotype of the disputed sire are shown for the 11 "StockMarks" markers in Table 4.4b. The matching probability for the 11 "StockMarks" markers of $3.9x10^{-11}$, is slightly better than for the 5 markers that were analysed manually. The identification of allele sizes using the ABI automated system is more precise than for manual interpretation of autoradiographs, as noted above. Nevertheless, if the same criteria for errors are applied and the probability recalculated using the pooled values for the observed next smallest and next largest alleles, the matching probability for the 11 markers becomes $1.85x10^{-5}$.

When the worst case is considered, the likelihood of finding an animal which is homozygous for the most frequent allele for each marker, for the 5 markers this would be $1.1x10^{-3}$, while for the 11 "StockMarks" this would be $2.6x10^{-5}$.

When DNA extracted from the straw from the disputed batch was compared with DNA from the 10 other undisputed collections from the same bull the genotype for all 16 markers matched exactly. The likelihood of finding a straw selected at random, which matches the genotype for all 16 markers when all errors are taken into account, would be $6.1x10^{-9}$. Thus the disputed straws are likely to contain the

TABLE 4.4 Probability of different genotypes in Charolais samples

A)

	HEL 6	CYP21	DRB3	FSHB	ETH131	All Markers
Bull genotype	252/254	267/271	153/161	186/186	131/147	
Probability of genotype using NWCBS data	0.0553	0.169	0.012	0.004	0.017	7.62×10^{-8}
Taking into account 2bp error	0.590	0.229	0.049	0.054	0.096	6.86×10^{-5}
Probability of genotype using UK data	0.0002	0.158	0.001	0.0001	0.001	1.57×10^{-15}
Taking into account 2bp error	0.007	0.368	0.016	0.005	0.020	4.12×10^{-9}

B)

	TGLA 227	TGLA 126	TGLA 122	TGLA 57	TGLA73	MGTG413
Bull Genotype	83/89	115/ 115	152/154	90/ 98	121/ 123	135/149
Probability of genotype	0.147	0.469	0.062	0.320	0.363	0.001
Taking into account 2bp error	0.323	0.507	0.506	0.428	1	0.085

	AGLA 293	TGLA 48	TGLA263	TGLA53	MGTG7	All markers
Bull Genotype	231/ 233	74/ 76	109/ 121	154/ 158	292/308	
Probability of genotype	0.427	0.256	0.217	0.021	0.155	3.92×10^{-11}
Taking into account 2bp error	1	1	0.247	0.062	0.204	1.83×10^{-5}

same semen as the 11 undisputed straws for that bull.

4.5.2.3 Discussion.

The alleles at an ideal microsatellite locus are expected to have a continuous distribution, with 2bp difference between alleles. Variations in allele frequency and alleles are found to differ between breeds. In addition some di-nucleotide microsatellite sequences are known to be flanked by runs of 10 or more single nucleotides, eg: FSHB used in this study (Fries et al. 1990). These microsatellite markers are known to have alleles which differ in size by only a single base pair. Determining a single base pair difference is difficult by manual analysis of autoradiographs, thus any single nucleotide variation was rounded to the next allele assuming a 2bp separation.

Using the 5 markers from the autoradiographic technique we find a greater chance of finding two animals with the same genotype when the allele frequencies from the NWCBS Charolais sires are used when compared with the samples from the rest of the UK. This is not surprising as the sires in one area are likely to be related, if only distantly, increasing the frequency of particular alleles in that area. In fact, in this study a brother of the test sire was included in the panel of 17 other sires. With the small sample sizes used it is possible that the observed allele frequencies are not representative of the breed, however the agreement between results for the NWCBS and other UK samples confirm the observed distribution of alleles is typical for the Charolais breed.

The 11 "StockMarks" markers give a similar, or slightly lower probability of two animals having the same genotype as the 5 other markers. One would expect a lower likelihood of two samples matching using 11 vs 5 markers, but two of the "StockMarks" markers have only three alleles (TGLA126, and AGLA293) in the 18 samples, while the lowest number of alleles is 6 for the other 5 markers (DRB3 and CYP21). In addition TGLA 126 has one allele at nearly 70% frequency, while none of the 5 manually interpreted markers has an allele at over 40% frequency (see table 4.3). Nevertheless, in this study the worst likely case with the 5 markers gave matching probabilities of at least 1 in 50,000, which is better than would be expected for Blood Typing where matching probabilities of 1 in 10,000 or less are expected

(Williams et al. in prep).

The number of markers used in these studies are clearly excessive, but does enable the comparison of different markers and methods of detection. Other than in the most exceptional circumstances about 10 markers with the levels of polymorphism seen here will be adequate to confirm the identity of an individual.

The results of this work were presented in court in defence of the AI centre accused of supplying incorrectly labelled semen straws. The evidence was accepted and the case against the centre dismissed.

The results of both cases cited above were successful examples of acceptance of evidence in court on the basis of DNA.

CHAPTER 5
ESTIMATION OF PHYLOGENETIC RELATIONSHIP BETWEEN BREEDS.

5.1 Introduction.

Archeological evidence suggests that farming spread in the neolithic age 9000 years ago in Europe (Medjugorac 1994). The change from food collection (hunting and gathering) to food production (agriculture and animal breeding) was followed by an increase in human population densities in areas where agriculture originated. The human populations, with their domesticated animals, expanded outward in all directions. This expansion in Europe took place at an average rate of one kilometer per year (Ammerman and Cavalli Sforza 1984). Cattle have had a central role in the evolution of human cultures and are the most economically important livestock species (Cunningham 1992).

The growing world population and expectations of higher living standards are steadily increasing. So is the demand for animal products, especially in developing countries. The dairy and beef industries have improved the quality of cattle products by improving the genetics of the animal itself. This usually involves attempts to change the genotypes of traditional breeds either by upgrading or more commonly by crossbreeding. In the modern world the process of replacing ancient native breeds with highly productive ones seems inevitable. Selection has had a great influence in the process of breed differentiation. Thus the different breeds of domestic cattle have developed through natural selection, artificial selection and genetic drift. A preliminary survey by FAO (1993) showed that domesticated cattle number over 1.2 billion on a global scale and consist of at least 780 breeds.

Because of their tremendous economic and cultural importance, much work has focused on the classification of cattle by assigning them to breed groups, but little is known about the early origin of domesticated cattle. The term "breed" is not uniformly applied in the classification of cattle populations, but describe the cattle population of a certain area with varying degrees of uniformity or expresses uniformity in a few traits affected by simple gene action such as colour, colour

pattern, head form, horn form, polledness etc. The breeds in turn are subdivided into populations, strains or lines. Different breeds have existed as socioeconomic entities for some hundreds of years.

Understanding the extent and pattern of genetic variability among cattle breeds may help in the development of more rational breeding programmes (Kidd et al. 1974) and is a prerequisite to develop conservation strategies for genetic resources (Kidd 1974). Furthermore, determination of the genetic relationships between cattle breeds should complement and clarify archeological data on the origin of animal domestication and husbandry.

The process of conservation of genetic diversity requires identification of genotypes (breeds) in danger of erosion or extinction. Estimating genetic distances between different breeds is useful to enable choice of those breeds that cover the widest range of genetic variability. Also the estimation of genetic distance helps to group breeds on the basis of geographical origin and their production characteristics: beef or dairy.

The best estimate of measuring genetic distance is obtained using genetically inherited polymorphism: blood groups, serum proteins and DNA markers. During the last few decades, blood group and protein polymorphisms have been used extensively for studying genetic distances in both man and cattle (Naik 1978), but these markers were not always useful for classifying closely related populations because the extent of variation was often quite low. However, these markers have been useful for estimating genetic relationships and the time that has elapsed since domestication of cattle (Medjugorac 1994).

The analysis of DNA sequences reveals the ultimate source of genetic variability, and thus offer highly informative markers for measuring genetic distances (Teale et al. 1994). Over the past decade, molecular genetic data have been used to address phylogenetic relationships. The D-loop region of mitochondrial DNA (mtDNA) has been found to be especially useful, because it is relatively short and highly variable (Avise et al. 1987), eg: this region of mtDNA is being used extensively for studying the genetic differentiation of human populations (Stoneking 1993). Phylogeny derived from mitochondrial sequence data provides sufficient

information to support a deep dichotomy between *Bos taurus* and *Bos indicus* (Loftus et al. 1994). To date a small number of population studies have examined mtDNA in cattle, focusing on local or regional breed groups (Watanabe et al. 1985, 1989, Bhat et al. 1990, Suzuki et al. 1993). Loftus et al. (1994) investigated the phylogenetic relationships between breeds chosen from Asia, Europe and Africa using RFLPs in mtDNA. Two major lineages were apparent: Afro-European and Asian. The lineages were found to differ between 0.5 to 2.3% in base sustitution in the mitochondrial genome and implies an ancient separation between *Bos indicus* and *Bos taurus* with an estimated divergence time of 200,000 to 1,000,000 years (Bradley et al. 1994) which was calculated from the estimate of 2-3% per million years (Brown et al. 1979). This suggests at least two independent domestications of cattle and the view is supported by microsatellite frequency data and Y chromosome polymorphisms (Cunningham et al. 1994).

Genetic variation in hypervariable loci has been used extensively for inter-population studies (Edwards et al. 1992, Pena et al. 1993). Microsatellite sequences have offered great potential for genetic comparisons within and between populations (Bowcock et al. 1994, Buchanan et al. 1994) and have facilitated the construction of phylogenetic trees of cattle that reflect their geographical origin or relatedness (Loftus et al. 1994, Mengers et al. 1994). Studies on microsatellite loci in cattle clearly show that they offer an excellent means of distinguishing closely related breeds.

Microsatellite loci have high mutation rates in comparison with other genetic markers (Weber and Wong 1993) and the average heterozygosity ranges between 30 and 80 percent. An understanding of the mechanism of mutations in microsatellite sequences is important if these markers are to be used in evolutionary genetics. Different views exist to explain the mutational mechanism of microsatellites: unequal exchange in meiosis (Baron et al. 1992) and slipped-strand mispairing during replication (Levinson & Gutman 1987, Buchanan et al. 1994). Edwards et al. (1992) and Valdes et al. (1993) have shown that the pattern of mutation for microsatellites follows a stepwise mutation model (Kimura and Ohta 1978). But according to another theory, slipped-strand mispairing (SSM) is accepted as the

predominant mutational event at microsatellite loci (Levinson and Gutman 1987) which follow an infinite isoallele model of mutation (Kimura and Crow 1964). This model assumes that new mutations are always different from the existing alleles in the population. Conversely, the stepwise mutation model implies the possibility of length convergence among alleles which are not identical by descent, a phenomenon which complicates comparison between populations. It is unclear which model for allelic distributions in microsatellites is correct. But the stepwise mutation model seems appropriate because mutations can cause an increase or decrease in repeat number (Valdes et al. 1993).

5.1.1 Measures of genetic distance.

A complete description of the genetic differences between any two breeds is not possible, but measures of genetic distance provide the best available objective description of their genetic differentiation. A choice has to be made about the markers used for measuring the distance, ie. whether to examine expressed loci with defined function or to address anonymous regions of the genome. It is ideal to use those markers which may not have been biased by selection. Clearly, it is necessary that the loci used display significant variation, and thus define as many differences as possible. The microsatellite loci used in earlier chapters have the potential for defining genetic variation.

Genetic distances are designed to express the variation between two populations as a single number (Smith 1977). The value gives a relative estimate of how distant the breeds are when compared to each other. If two population have the same allele frequency distribution at the same locus the genetic distance between them at least for that locus, is zero. If the populations have no alleles in common at any locus the distance may be separated by the maximum value which can be infinite (Swofford and Olsen 1990). Genetic distances may be regarded as a means of comparing pairs of extant populations or as a basis for constructing evolutionary histories for the population.

Many measures for genetic distances have been proposed (Edwards and Cavalli-Sforza 1964, Kimura 1968, Balakrishnan and Sanghvi 1968, Nei 1972,

Reynolds 1983, Nei 1987) based on allele frequency data. Edwards and Cavalli-Sforza (1963, 1964) made the first attempt to use gene frequencies to relate phylogenetic relationships. When the allele frequency data for many loci are available the genetic distance estimate is obtained by averaging over these loci. Genetic distance analysis allows a ranking of breeds and populations according to their levels of phylogenetic distinction.

Genetic distances are used either for estimating evolutionary times or for constructing phylogenies (dendrograms). A distance measure that is appropriate for estimating evolutionary time is not good for obtaining the correct topology (Nei and Takezaki 1994). Even though all the methods of calculation of genetic distances are based on allele frequencies, the assumptions made are different for each method.

Nei's genetic distance (Nei 1972) is formulated for an infinite isoallele model of mutation, in which each mutant forms a completely new allele. It is assumed that all loci have the same rate of mutation, and that the genetic variability initially in the population is at equilibrium between mutation and genetic drift. But Cavalli Sforza's chord measure (Cavalli Sforza and Edwards 1967) and Reynold et al. (1983) genetic distance assume that there is no mutation, and that all gene frequency changes are by genetic drift alone. These measures of genetic distances are different estimators of the same quantity under the same model and are based on the assumption that the differences between population arise from genetic drift.

It is clear that different approaches have provided different answers at least in the best studied livestock species, cattle. It is apparent that Y chromosomes examined in some population of *Bos taurus* breeds in Africa have been derived from those of *Bos indicus* subspecies (Bradley et al. 1994). Conversely, analysis of mtDNA did not reveal *Bos indicus* having mixed with African *Bos taurus* populations (Suzuki et al. 1993). In comparing African zebu and taurine breeds, microsatellite analysis provides yet another view of the relationship between these types. The relationship seems intermediate between that resulting from the Y chromosome and mtDNA polymorphism studies (Teale et al 1994).

5.1.2 Phylogenetic trees.

Phylogenetic trees are graphical representation consisting of nodes (taxonomic units) and branches (pathway connecting nodes) that summarize the evolutionary relationships among breeds. There are many statistical methods for constructing phylogenetic trees or dendrograms from molecular data. They can be classified into distance methods and discrete character methods. Distance matrix methods are based on the set of distances calculated between each pair of taxonomic units. The configuration of the resulting tree depends on the quality of the distance measure. Popular distance methods include unweighted pair group method with arithmetic means (UPGMA) (Sneath and Sokal 1973) and the neighbor-joining (NJ) approach later developed by Saitou and Imanishi (1989) and Rzhetsky and Nei (1992). UPGMA depends on the assumption that the rate of evolution is the same for all evolutionary lineages, so that any true heterogenity would remain undetected. But the NJ method allows unequal rates of evolution, thus eliminating such risks. In the discrete character method, a tree is constructed by considering the evolutionary relationships of DNA sequences at each nucleotide site. Two major methods belonging to this category are the maximum parsimony (MP) method (Eck and Dayhoff 1966) and the maximum likelihood (ML) method (Felsenstein 1981).

The phylogenetic trees may be represented either as rooted or unrooted trees. A rooted tree conveys the options of ordering of the species on a tree, while an unrooted tree merely reflects distances between units with no notion of ancestral relationship.

The accuracy of the tree is dependent on the number of loci typed, nature of the samples used, the treemaking method and the distance measure used for analysis. The species under study also has an effect. The populations that diverged for millions of years are to be regarded as different from those that diverged recently. The species which are widely separated requires few markers for accurate comparison in contrast to the closely related species. Nei et al. (1983) have shown that even with 30 loci the constructed tree could be incorrect. Later Nei and Takezaki (1994) worked out the confidence level of contructing trees using different number of loci by simulation. The percentage of replications for the correct tree

topology varied between 82-96% with the different methods of genetic distance even with 100 loci, which suggests that if the true genetic distances are less than 0.004 and if there are many branches, the tree would be incorrect even with 100 markers. However the calculation of genetic distance is dependent on the number and frequency of alleles of the markers used for calculating the genetic distance.

The current intensive characterisation of microsatellite polymorphism in a wide range of species suggests that these markers are very useful in population and ecological genetic studies. The aim of this chapter is to use the data collected for analysis of microsatellite loci in parentage verification to test the methods of measuring genetic relationships in cattle breeds based on allele frequencies.

5.2 MATERIALS AND METHODS.

The genetic relationship between fifteen breeds of cattle was assessed with the microsatellite markers detailed in Table 3.3.

5.2.1 *Phylogenetic Analysis.*

The computation of genetic distance and the phylogenetic analysis were done using the programs in the PHYLIP package (Felsenstein 1993). Cavalli Sforza's chord measure (Cavalli sforza and Edwards 1967), Nei's (1972) and Reynold's (Reynold et al. 1983) genetic distance methods were calculated. Dendrograms were constructed by the neighbor joining method. Bootstrap resampling (Felsenstein 1985) using 1000 replicates of the loci from the original data set was used to test the robustness of the relationships established. The programs used in the analysis were:

1. *Gendist:* This program computes the genetic distance matrix from a set of gene frequencies in different populations (or species). Nei's genetic distance (1972) based on infinite isoallele mutation model was calculated using the formula

$$D = -\ln \frac{\sum_m \sum_i p_{1mi} p_{2mi}}{\left[\sqrt{\sum_m \sum_i p_{1mi}^2}\right]\left[\sqrt{\sum_m \sum_i p_{2mi}^2}\right]}$$

where m is summed over loci, i over alleles at the m^{th} locus and p_{1mi} and p_{2mi} are the

frequencies of the i^{th} allele at the m^{th} locus in populations 1 and 2 respectively. Cavalli Sforza's chord distance is given by

$$D^2 = \frac{4\sum_m \left[1 - \sum_i \sqrt{p_{imi}}\sqrt{p_{2mi}}\right]}{\sum_m (a-1)}$$

where m indexes the loci, i is summed over the alleles of the m^{th} locus, a is the number of alleles at the m^{th} locus and D^2 is the quantity that is expected to rise linearly with cumulated genetic drift.

Reynold's genetic distance was calculated using the formula

$$D^2 = \frac{\sum_m \sum_i \left[p_{1mi} - p_{2mi}^2\right]}{2\sum_m \left[1 - \sum_i p_{1mi}p_{2mi}\right]}$$

Notations are the same as above.

2. *Neighbor:* This program constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join to select the "best" or "close to the best" tree (Saitou and Nei 1987). The tree is not rearranged thereafter. The tree does not assume an evolutionary clock, so that in effect it is an unrooted tree.

3. *Drawgram:* This program interactively plots a phenogram-like rooted tree diagram.

4. *Seqboot:* This is a general bootstrapping tool. It is intended to generate a multiple data set that are resampled versions of the input data set. Bootstrap resampling using 1000 replicates of the loci from the original data set was used to assess confidence in the branching order of the phylogeny.

The use of bootstrapping involved in seqboot analysis and its use in phylogeny estimation was introduced by Felsenstein (1985). It involves creating a new data set by sampling N characters randomly with replacement, so that the resulting data set has the same size as the original in such a way that every one of the original observations has an equal chance of being chosen at any stage. Some

of the original sample elements will not, therefore, appear in any bootstrap sample while some may appear many times. The random variation of the results from analysing the bootstrapped data set can be shown statistically to be typical of the variation shown from collecting new data sets. The method assumes that the characters evolve independently, an assumption that may not be realistic for many kinds of data.

5.3 RESULTS.

Allele frequencies of the five microsatellite loci against 15 breeds included in this study are shown in Table 3.5 - 3.9 . Certain breeds had some alleles occurring at a higher frequency than others. Also, for a given marker the allele with the highest frequency in one breed may be absent in others. In general, all microsatellites showed a large variation in allele frequency within and between breeds. Such an uneven distribution of allele frequencies among breeds is true for most of the typed markers.

The genetic distance estimates using the three methods - Nei's (1972), Cavalli Sforza's chord measure and Reynold's genetic distance are given in Table 5.1 - 5.3. In these three methods of calculation of genetic distance, the smallest value was found between FR and LN indicating a greater similarity between them than other pairs of breeds. Similarly a small genetic distance was observed between JE and GU in each of these methods of calculations. The largest distance was found between BO and BG. In the three genetic distance methods, BO had a larger genetic distance with other breeds. This was not surprising considering the geographical origin of both breeds.

The phenogram obtained from all the three genetic distances using the neighbor joining method is shown in fig 5.1 - 5.3. Also the neighbor joining tree shows the bootstrap values over 1,000 replications of the resampled loci. The BO was used as an outgroup in all the methods for relating other breeds in view of its differences in geographical location and the morphological appearance from other breeds.

In the phenogram constructed using Nei's genetic distance the tree consists

Table 5.1 Genetic distance between fifteen breeds of cattle based on Nei's method.

AA													
AY	0.5399												
BB	1.0378	0.7647											
BA	1.0595	0.8812	0.9023										
DX	1.2114	0.8757	0.9001	1.1779									
CH	1.1260	1.0608	0.5990	0.9757	0.8739								
FR	0.8680	0.6935	0.5783	0.5243	0.9126	0.6645							
GU	0.8138	0.7450	0.6311	0.9071	1.1754	0.6028	0.6872						
HF	1.3991	0.5852	0.7925	1.0938	0.7967	1.0204	1.0383	0.9877					
JE	1.1872	0.7786	0.6694	0.8408	0.8416	0.8710	0.6172	0.4666	1.2560				
LN	1.1533	1.0347	0.7724	0.5800	1.0559	0.8617	0.4271	0.8063	0.8604	0.6696			
SG	1.2914	0.6826	0.5766	0.5874	1.0409	0.5440	0.7054	0.6736	0.4939	0.6371	0.5220		
BG	1.6341	1.3282	0.6810	1.0795	1.5134	1.4044	0.7947	1.4106	1.2258	0.9535	0.8559	0.8939	
SD	1.0049	0.9970	0.7150	0.9971	1.1916	1.4870	0.6789	1.1939	0.9565	0.9942	0.9001	0.9604	0.5063
BO	1.8396	1.3948	1.6906	1.4989	1.4199	1.2803	1.3975	1.5019	1.4034	2.0507	1.4575	1.3428	2.3633	1.7107

Table 5.2 Genetic distance between fifteen breeds of cattle based on Cavalli's method.

AA

AY 0.0781

BB 0.0816 0.0778

BA 0.1050 0.0873 0.0840

DX 0.1151 0.0967 0.0930 0.1188

CH 0.0997 0.1055 0.0776 0.0961 0.0992

FR 0.0889 0.0834 0.0639 0.0694 0.1111 0.0787

GU 0.0837 0.0836 0.0800 0.0917 0.1138 0.0715 0.0773

HF 0.1104 0.0550 0.0830 0.1056 0.0788 0.0914 0.0857 0.0827

JE 0.1096 0.0858 0.0811 0.0960 0.0935 0.0906 0.0707 0.0647 0.1008

LN 0.1057 0.0886 0.0739 0.0682 0.1068 0.0869 0.0558 0.0800 0.0884 0.0833

SG 0.1087 0.0787 0.0638 0.0712 0.0974 0.0754 0.0758 0.0715 0.0693 0.0787 0.0659

BG 0.1196 0.1196 0.1053 0.1187 0.1308 0.1128 0.1047 0.1185 0.1106 0.1136 0.0980 0.0937

SD 0.1010 0.0945 0.0975 0.0955 0.1141 0.1227 0.0829 0.1067 0.0951 0.1074 0.0976 0.0952 0.0916

BO 0.1432 0.1248 0.1277 0.1353 0.1249 0.1249 0.1187 0.1176 0.1269 0.1480 0.1199 0.1217 0.1495 0.1300

Table 5.3 Genetic distance between fifteen breeds of cattle based on Reynold's method.

AA

AY 0.0906

BB 0.1340 0.1045

BA 0.1465 0.1238 0.1260

DX 0.1643 0.1312 0.1337 0.1637

CH 0.1455 0.1310 0.0946 0.1370 0.1369

FR 0.1070 0.0847 0.0759 0.0801 0.1195 0.0884

GU 0.1267 0.1117 0.1013 0.1358 0.1626 0.1027 0.0940

HF 0.1598 0.0930 0.1129 0.1452 0.1304 0.1353 0.1136 0.1372

JE 0.1453 0.1078 0.0988 0.1234 0.1311 0.1211 0.0814 0.0840 0.1452

LN 0.1367 0.1192 0.1021 0.0938 0.1409 0.1144 0.0586 0.1140 0.1146 0.0956

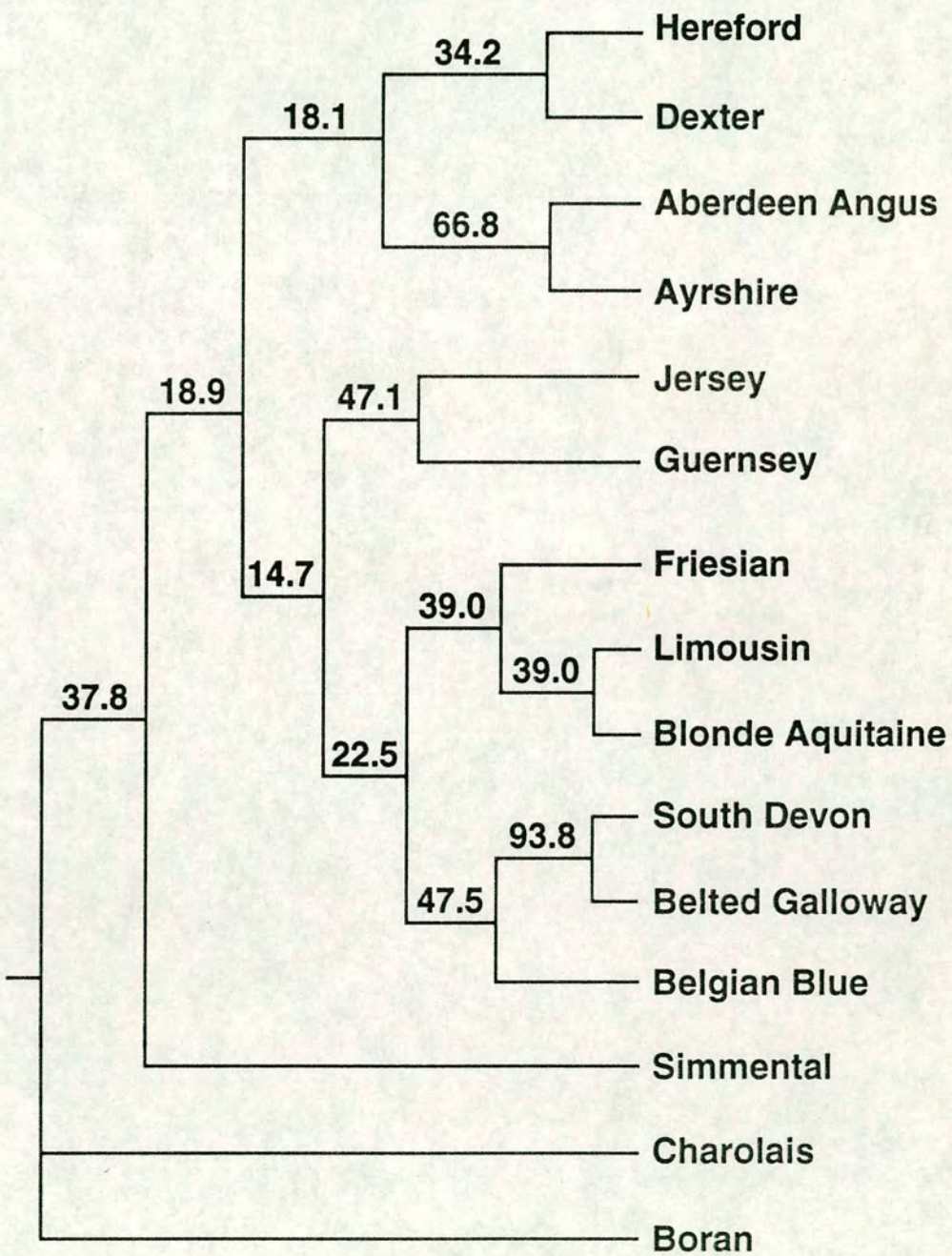
SG 0.1464 0.0962 0.0867 0.0964 0.1425 0.0877 0.0849 0.1044 0.0822 0.0946 0.0781

BG 0.2433 0.2169 0.1624 0.2116 0.2481 0.2281 0.1638 0.2328 0.2177 0.1931 0.1787 0.1843

SD 0.1448 0.1342 0.1121 0.1455 0.1667 0.1670 0.0961 0.1573 0.1379 0.1370 0.1241 0.1309 0.1443

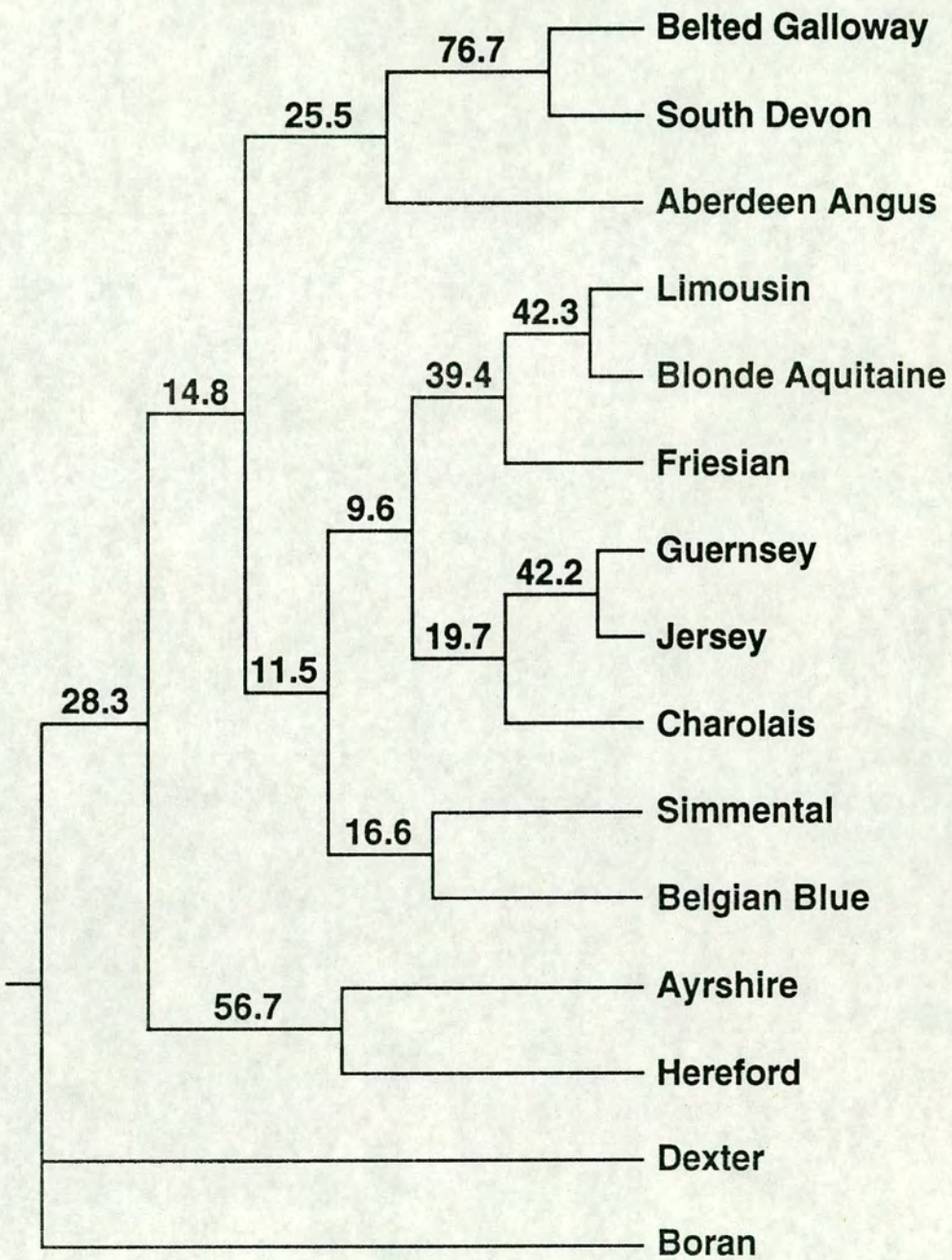
BO 0.2367 0.2067 0.2202 0.2246 0.2293 0.2073 0.1891 0.2237 0.2145 0.2342 0.2060 0.2031 0.3237 0.2361

Figure 5.1 Dendrogram constructed by neighbor joining method based on Nei's distance



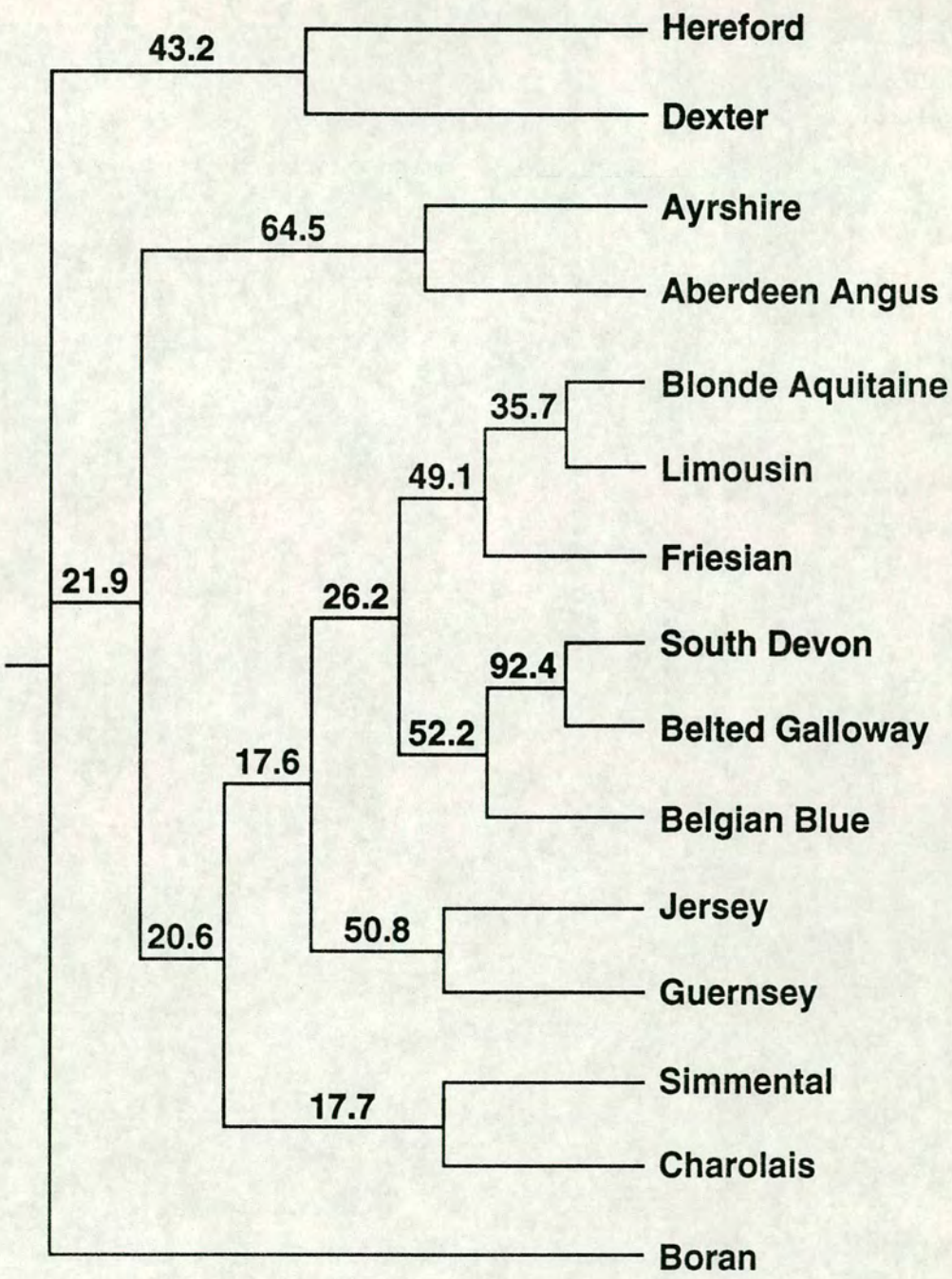
Numbering on the nodes indicates the number of times a particular branch was recovered (in percentage) following 1000 bootstrap replications.

Figure 5.2 Dendrogram constructed by neighbor joining method based on Cavalli's distance



Numbering on the nodes indicates the number of times a particular branch was recovered (in percentage) following 1000 bootstrap replications.

Figure 5.3 Dendrogram constructed by neighbor joining method based on Reynold's distance



Numbering on the nodes indicates the number of times a particular branch was recovered (in percentage) following 1000 bootstrap replications.

of two major groups: BO and CH forming separate branches of their own and the other breeds forming a separate group with five subgroups. This indicates that BO and CH are significantly different from the other breeds. 94% of the 1000 bootstrap replicates gave a SD/BG node and 67% supported the node at the AA/AY branch. The most important features of the topology are the pairing of the Scottish breeds - AA and AY, the Channel Island breeds JE and GU, and the French breeds LN and BA. CH is separated from all the other breeds. HF and DX are seen clustered together with AA and AY.

In the phenogram constructed with Cavallis chord distance, BO and DX were clustered as two distinct out-groups and all other breeds formed another group with five subgroups. 77% of 1,000 replicates supported the node at SD/BG and 57% at AY/HF node. The tree topology was different from that drawn with Nei's genetic distance. The Channel Island breeds JE and GU were grouped with the French breed CH, and BG and SD with AA. SG was grouped with BB and AY with HF.

The phenogram drawn with Reynold's genetic distance showed only BO as a distinct outgroup, the other breeds forming six subgroups. 65% of the bootstrapped values showed an AA/AY node, 93% for SD/BG and 51% for JE/GU node. The tree topology was similar to that drawn by Nei's distance. The pairing of all the groups and subgroups was consistent except for CH/SG.

To test the effect of each microsatellite on the tree topology, phenograms were constructed with the data from four microsatellites each time, excluding the fifth and using all the three genetic distance methods. Trees showed the consistent pairing of SD & BG and JE & GU and FR, LN & BA. Phenograms were also constructed for all breeds for each microsatellite marker separately. No consistent topology was seen between any of these phenograms.

In the pooled data with all the five markers the overall topology of the trees drawn with the three distance matrices was different from each other. The pairing of SD & BG, JE & GU and FR, LN & BA were consistent. The low bootstrap values between breeds suggests either a historical admixture or a recent divergence between the populations. It is likely that the samples collected were not representative of the whole population while, the number of loci typed are likely not

to be a full representation of the entire genome.

To compare the correlation between the three distance methods, the genetic distance values between breeds were plotted against each other as a scatter diagram (fig 5.4 - 5.6). The graphs plotted had a linear pattern, showing that the three methods are similar to one another. Each plot on one graph has almost a similar position in the other. For example, the breeds with larger genetic distance were plotted at a similar position in all the three graphs. The reverse was also true.

5.4 DISCUSSION

All modern domesticated cattle breeds (zebu and taurine) are believed to have been derived from the now extinct wild ox or aurochs, *Bos primigenius* (Payne 1991). Large scale cattle population movements following human migrations, population growth, genetic drift and different proportion of aurochs genes are the different factors that influenced the initial differentiation of cattle populations (Medjugorac et al. 1994). The effect of domestication, selection based on different traits and also the geographical location led to the development of different breeds which would give a possible explanation for the variation in genetic distances between the 15 breeds examined in this study.

Even though many genetic markers have been suggested for calculation of genetic distances, the highly polymorphic microsatellite markers are better indicators of the evolutionary relationships within a particular species (Buchanan et al. 1994, MacHugh et al. 1994). The two important features of microsatellite loci - the mutation rate and the conservation between breeds make them a good measure for phylogenetic studies. Relatively few microsatellite mutations have been observed to date. An upper limit estimate of the microsatellite mutation rate in recombinant inbred strains of mouse was reported to be in the range of 10^{-2} and 10^{-4} (Dallas 1992). Kwiatkowski et al (1992) observed 4.5 X 10^{-4} spontaneous mutations per locus per gamete in humans. Some microsatellites are highly conserved between species. Maichele and Chamberlain (1992) bridged a huge evolutionary distance by amplifying a dinucleotide repeat associated with the dystrophin gene in both man and the mouse. Microsatellites are not good markers for genetic comparison between

Fig 5.4 Comparison of Nei's and Cavalli's genetic distance methods.

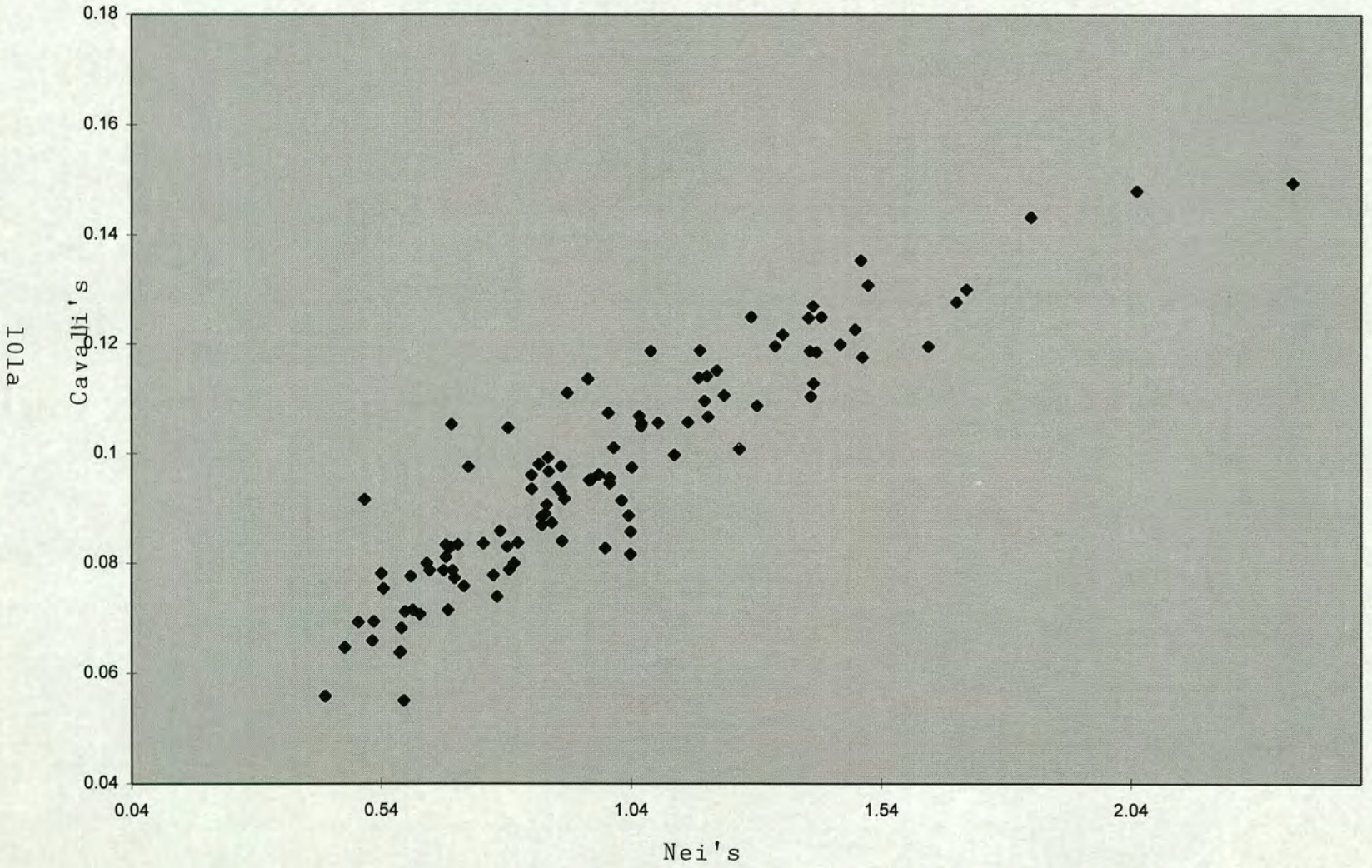


Fig 5.5 Comparison of Nei's and Reynold's genetic distance methods.

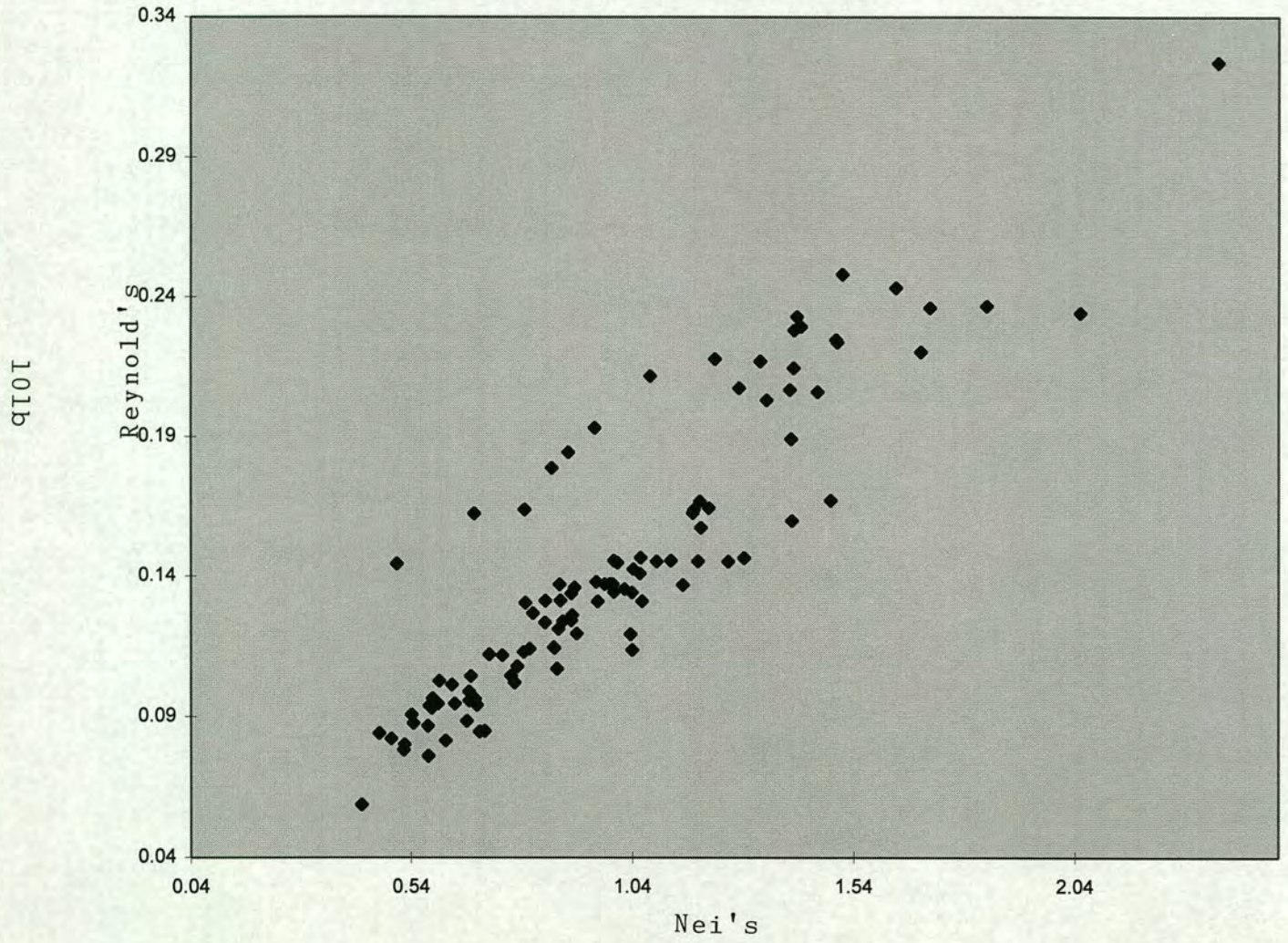
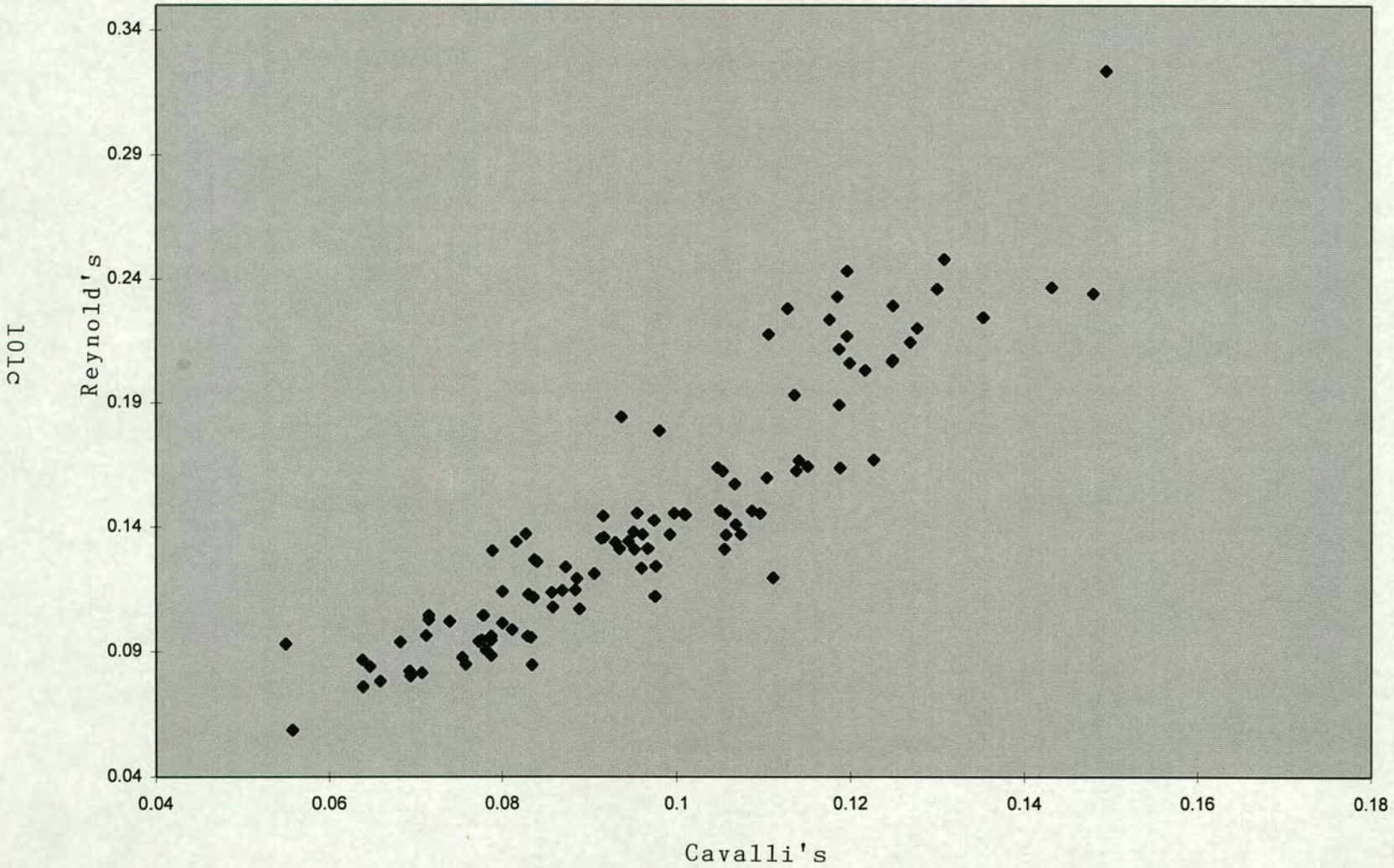


Fig 5.6 Comparison of Cavalli's and Reynold's genetic distance methods.



species. The nature of mutation in microsatellite loci is such that a limit is set for the expansion for the number of repeats. Because of this limitation genetic variation expected between species cannot be detected using microsatellites.

Significant variation in allele frequency was noted between the 15 breeds with the five microsatellite loci studied. For some breeds like DX, BG & BO only a few alleles were seen for all the markers when compared to others like FR & LN. This was an indication of the degree of genetic variability within breeds. The variation in allele frequency in different microsatellite loci could be effectively used to identify breeds if the samples are representative of the breeds.

Different methods of calculation of genetic distance have been described and the merits of each have been reviewed by Felsenstein (1985). It has been demonstrated that the allelic distribution of microsatellite loci are consistent with the mutation model in which a direct relationship exists between allelic length difference and evolutionary divergence (Valdes et al. 1993). Nei's genetic distance (1972) based on microsatellite allele frequency (MacHugh et al. 1994), Reynold's weighted average distance (DW) based on blood groups related allele frequency (Medjugorac et al. 1994) and Cavalli's chord distance based on the variation in mitochondrial DNA sequence (Loftus et al. 1994) have been used to assess breed relationships in cattle. Similarly different approaches (neighbor joining, maximum parsimony, maximum likelihood and UPGMA) have been made for constructing evolutionary trees (Nei and Takezaki 1994). Penny et al. (1992) reported that neighbor joining method for distance data as the most effective because the method requires less computer time. The program requires a time proportional to the square of the number of species. Moreover, neighbor joining is well suited to bootstrapping studies and for analysis of very large trees. In the present study the three methods of calculations of genetic distance included in the PHYLIP package - Nei's genetic distance, Cavalli Sforza's chord distance and Reynold's genetic distance were evaluated with the allele frequency data for five microsatellite markers in 15 breeds of cattle. All the trees were drawn using the neighbor joining method, because the option for setting an outgroup was not available using UPGMA method.

Trees were drawn for each marker separately for all the five markers and

also by eliminating one marker each time for all the methods of genetic distance to determine whether there was any significant effect on any of the markers. No definite conclusion was drawn except for a consistent pairing of JE and GU, and SD and BG in all the three topologies. This may be because in these breeds only few sires were used for breeding and thus genetic variation in the population is less. The subgrouping between breeds was similar in the topologies constructed using Nei's and the Reynold's methods, even though these subgroups were grouped differently. The clustering of breeds based on Cavalli's chord distance was very different from that of Nei's and Reynold's. All the trees were drawn based on bootstrap values, and the tree topology was consistent for each method even when bootstrapped again.

The estimated distance values and groupings summarised in the phenogram constructed on Nei's distance can be compared with the classification scheme discussed by Baker and Manwell (1991) based on protein polymorphism. In their study, HF and AA breeds were clustered as expected on biogeographical grounds. However, in this study, even though these two breeds were grouped together, HF was subgrouped with DX, and AA with AY, resulting in mixing the English, Scottish and Irish breeds. AA is a highly developed beef breed whose ancestry may be traced to the 19th century when the AY breed was used to improve this breed (Felius 1985). Thus a smaller genetic distance between these breeds may be expected. The Channel Island breeds JE and GU were grouped together, and this was not surprising based on the geographical proximity of their origin. LN and BA, the French breeds were in the same cluster, but grouped together with FR which is also a mid European breed. SD and BG were grouped together with BB all of which are beef breeds, although from apparently different geographical origin. This lack of differentiation among these breeds may be due to the historical mixing for upgrading programmes before the demarcation of various groups into reproductively isolated breeds in the 19th century (MacHugh et al. 1994) or could be purely by chance. The same subgrouping were also seen on the phenogram based Reynold's distance. The status of the CH breed forming a distinct group from other breeds is in accordance with the migration of cattle to France by the Romans (MacHugh et al. 1994). In the phenogram based on Reynold's distance, CH was

grouped with SG, a Swiss breed, and in Cavalli's method CH was grouped with JE and GU among all other breeds. The position of CH seems to be somewhat arbitary. Even though the BO is clearly an outgroup, the present day BO is thought to have a taurine origin (Felius 1985) by using Taurine bulls on Indicus females.

Constructing intraspecies phylogenies is difficult for species like cattle because of large scale population movements and the known hybridisation of populations (Crandall and Templeton 1993). A better picture of phylogeny may be obtained if the markers chosen for analysis are a representation of the whole genome and the samples chosen are that of the population.

It is seen from the present study that none of the three methods of calculation of genetic distance gave conclusive results. For a precise phylogenetic analysis, sample size analysed is not of great importance provided the samples are a representative of the population. Nei and Roychoudhury (1972) concluded that for estimating genetic distance, a large number of loci rather than a large number of individuals per locus has to be examined.

Comparison of the phylogenetic trees on the same sample of animals using allele frequencies from blood groups with that of microsatellites would clarify if functional and neutral loci have any effect on the tree topology. Similarly sampling different breeds based on the geographical origin and constructing phylogenetic trees with allele frequencies on microsatellite loci included in this study would show the effect of sampling on tree topology.

Nei and Takezaki (1994) have generated allele frequency data for 100 microsatellite loci for eight populations and computed genetic distance using nine different methods for the first 10,20,30 .....and 100 loci separately. The phylogenetic tree constructed by UPGMA and NJ methods and the percentage of replication in which the correct topology was obtained was computed for each data set and distance measure. It shows that for Nei's (1972) method only 71% of replicates gave the correct topology when allele frequency from 100 microsatellite loci (heterozygosity 0.5) were used. But in the present study the clustering of breeds based on Nei's distance was in agreement with the classification of beef and dairy populations and also it reflected on the geographical origin of breeds. Even

though the samples size used was small, the clustering was in agreement with the findings of Baker and Manwell (1991). However, it is possible that this is by chance.

Microsatellites are sufficiently stable for informative comparisons between breeds and therefore can be successfully used in ecological and population genetics. The phylogenetic tree constructed using microsatellites identifies the breeds having the largest amount of genetic distance from others which therefore need to be preserved and thus helps conservation of genetic resources.

CHAPTER 6

ATTEMPTS TO DETECT THE MUTATION CAUSING DWARFISM IN DEXTER CATTLE.

6.1 Introduction

6.1.1 Characteristics of Dexter cattle.

The Dexter breed of cattle originated in South Western Ireland. Historically, the breed was developed by Mr. Dexter in 1750's from amongst the best of the hardy mountain cattle of the area. In 1882, these cattle were first introduced into England. The Dexter is the smallest breed of British cattle, the average weight of a cow being 350kg and a height of approximately 0.9 - 0.95m at the shoulder. Although usually black in hue, dark red and dun Dexters are also common. The colour is usually solid with a few white markings on the udder and in the region behind the navel. Horns are rather small and stumpy. Being a mountain breed, the animals are reputedly hardy and efficient food converters. The breed is a typical dual purpose type (milk + beef) and is distributed across many parts of the world. Show type Dexters are selected for having very short legs, which the breed carry as a genetic defect that results in still births.

6.1.2 The Dexter dwarf phenotype.

Several types of dwarfism are known. Dexter cattle are classified as disproportionate dwarfs (chondrodystrophia), which is differentiated from primordial or proportionate dwarfism. Proportionate dwarfism is a recessive condition which is fixed by selection. Similar selection procedures have resulted in the development of specific small breeds such as the miniature Poodle dog. Proportionate dwarfism in man, also known as hypopituitary dwarfism, is caused by reduced growth hormone levels (Erlebachar et al. 1995) and the individuals are phenotypically small.

In contrast, Dexter cattle are a classical example of disproportionate dwarfism. The heterozygote phenotype of Dexter is characterised by a broad wide head (brachycephaly), deep and well ribbed body with legs noticably short from knee to fetlock (micromelia), large hoofs and a tail seeming to originate forward along the back. Homozygous normal Dexter cattle are long-legged in comparison with

heterozygous carrier animals. A comparison between homozygous long-legged and heterozygous short-legged Dexter cattle, and the comparative size of a short-legged Dexter with a normal bull are shown in Fig 6.1a and 6.1b. In the radiographs of Dexter carriers, the frequently recognised diagnostic features are: longitudinal compression of the bodies of the last four or five thoracic and all of the lumbar vertebrae, and in particular the convex "concertina like" undulations on the ventral aspect of the bodies of these vertebrae (Weaver 1975). Not all Dexter carriers show such marked features, and the less extreme forms are easily mistaken as individuals of normal phenotype. This chondrodystrophy accounts for 12.6% of all congenital defects of the skeletal system in bovine (Greene et al. 1973).

Dexter dwarfism is a homozygous lethal condition and is dominantly inherited with 100% penetrance, with still born offspring (Weaver 1975). Homozygous Dexter cattle or the so-called "bull-dog" calves are often aborted before the eighth month of gestation. The characteristic features of a Dexter "bull-dog" calf are shown in Fig 6.2. These calves have phocomelic or very short legs, so that the hooves appear to project directly from the body. The cranium is vaulted, and the pre-maxillary region of the head shortened so that the tongue protrudes. The body tissues are oedematous and the tail seems to arise from far behind the back. Umbilical herniation and a large amphalocoele are also commonly observed (Crew 1924).

6.1.3 The origin and frequency of the Dexter dwarf gene.

The Dexter mutation is believed to have originated in the larger Irish Kerry breed. This defect was then selected to high frequency to produce the Dexter breed of cattle which is now established in the UK, USA, Canada and Australia. In the UK, between 1984 and 1988, 3967 calves were born in Dexter herds, predominantly from matings between carriers and non-carriers. During this period 119 "bull-dog" calves and 66 abortuses of unknown phenotype were reported (The Dexter Cattle Society Bulletin 1994).

Dominantly inherited dwarfism has also been reported in many other breeds of cattle, especially in Hereford and Friesian, but also in the Aberdeen Angus, Ayrshire, Shorthorn, Charolais, Guernsey, Jersey and Romagnola as well as in some cross-breeds

Figure 6.1a Comparison between homozygous long-legged and heterozygous short-legged Dexter cows

Figure 6.1b Comparison between a heterozygous short-legged Dexter cow and a Limousin bull

Figure 6.2 Phenotype of a Dexter "bull-dog" calf

(Weaver 1975). "Bull-dog" calves are also reported to have been born in some of these breeds, however at a much lower frequency than in Dexter herds. Moreover, in Herefords and in some other breeds, the homozygous calf is live-born and is known as a "snorter" because of its distinctive snout. Presumably, a different gene or a modifying gene or a different type of mutation within the same gene is associated with dwarfism in these other breeds. As in cattle, a similar achondroplastic condition is seen in so called Creeper fowl (Hutt 1958). The creepers, are also heterozygous for a dominant gene which is lethal when homozygous. As a result, a quarter of the developing embryos from two creeper parents die, usually at 3-4 days of incubation.

Dwarfism in Dexter cattle is considered to be the bovine equivalent of the dwarfism - achondroplasia (ACH) in man. Dominantly inherited ACH is the commonest genetic form of dwarfism in man, with an estimated frequency of 1 in 26,000 and with at least 80% of the cases being sporadic (Oberklaid et al. 1979) indicating that the condition is caused by a gene with a high mutation rate. The heterozygote phenotype of the classical ACH in man is characterised by disproportionate short stature with proximal shortening of the extremities, macrocephaly, frontal and biparietal prominence and exaggerated lumbar lordosis (Jones 1988). Homozygotes for the ACH gene suffer from a severe disorder of the skeleton, but these have radiological changes different from those of heterozygotic carriers of ACH. Homozygote ACH in man does not usually result in still births, with some children surviving beyond infancy (Pauli et al. 1983, Stanescu et al. 1990). Most homozygote ACH cases do, however, face premature death either from respiratory failure due to the small thoracic cage or neurological deficit from hydrocephalus.

A gene causing ACH in man has been mapped to the chromosome 4p16.3 (Velinov et al. 1994, Le Merrer et al. 1994) and has been shown to be caused by point mutations in the transmembrane (TM) region of FGFR3 (Shiang et al. 1994, Superti-Furga et al. 1995).

6.1.4 The role of Fibroblast growth factors and growth factor receptors in development.

Fibroblast growth factors (FGFs) constitute a family of closely related

polypeptide mitogens (Folkman and Klagsbrun 1987, Thomas 1987, Burgess and Maciag 1989). Currently, nine members of this gene family have been identified on the basis of amino acid sequence homologies (Givol and Yayon 1992). These proteins are highly conserved across species. Bovine acidic FGF (aFGF) and basic FGF (bFGF) differ from those of man by only 12 and 2 amino acids respectively (Thomas 1987). Various functions associated with different FGFs are reviewed by Muenke and Schell (1995) (Table 6.1). FGF2 is shown to be the prime candidate for the chick limb bud apical ridge growth signal (Fallon et al. 1994). The role of FGFs in branching morphogenesis of the mammalian lung has also been demonstrated (Peters et al. 1994). The ability of FGFs to influence cellular differentiation as well as expression during embryogenesis suggests that FGFs play an important role in development.

Fibroblast growth factors use a dual receptor system to activate signal transduction pathways (Klagsbrun and Baird 1991, Rapraeger et al. 1991). The primary component of this system is a group of signal transducing fibroblast growth factor receptors (FGFRs). A family of four different but highly homologous human FGFR genes have been identified - FGFR1, FGFR2, FGFR3 and FGFR4, and have been mapped to chromosomes 8, 10, 4 and 5 respectively in man (Oldridge et al. 1995). The general schematic representation of fibroblast growth factor receptors is shown in Fig 6.3a. The receptor protein has a membrane spanning region, an amino terminal signal peptide and three extracellular immunoglobulin like ligand domains (Williams and Barclay 1988) and an intracellular tyrosine kinase domain (Basilico and Moscatelli 1992, Johnson and Williams 1993). Between the first and the second ligand binding domain, there are eight consecutive acidic residues referred to as the "acidic box". The intracellular tyrosine kinase domain contains a cysteine rich region (Burrus et al. 1992). The second component of this receptor consists of heparan sulphate proteoglycans or selected heparin like molecules that are required in the order for FGFs to bind to and activate the FGFRs (Yayon et al. 1991, Moscatelli 1987). Although the mechanism by which heparin/heparan sulphate activates FGF is unknown, heparin, FGF and the FGFR can form a trimolecular complex (Ornitz and Leder 1992). Heparin or heparan sulphate may interact directly with the FGFR linking it to FGF (Ornitz et al. 1995). Upon ligand binding, the FGFRs are known to dimerize prior to intracellular signal

Table 6.1 Functions of the fibroblast growth factor (FGF) family.

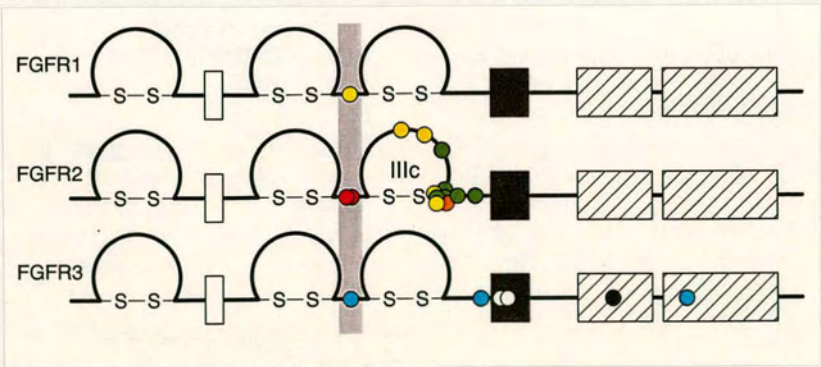
FGF	Gene or protein pseudonyms	Associated functions
FGF1	Acidic fibroblast growth factor Endothelial cell growth factor	Endothelial cell migration and proliferation, angiogenesis, delayed-early activation.
FGF2	Basic fibroblast growth factor	Angiogenesis.
FGF3	Oncogene *int2*	Formation of the inner ear spatial patterning processes. Mammary carcinoma.
FGF4	Oncogene *bst*	Human stomach cancer, melanoma, Limb development.
FGF5	Oncogene fibroblast growth factor	Inhibitor of hair elongation.
FGF6	Oncogene *bst2*	_
FGF7	Keratinocyte growth factor (KGF)	Epidermal growth and wound healing, branching morphogenesis of lung, salivary gland and prostate gland.
FGF8	Androgen induced growth factor (AIGF)	Androgen dependent tumor cell proliferation.
FGF9	Glia activating factor	Glial cell proliferation.

(Reference: Muenke and Schell 1995)

Figure 6.3a Schematic representation of FGF receptors

S-signal sequence; Ig-immunoglobulin like domains (I, II and III); A-acidic region; TM-transmembrane domain; TK-tyrosine kinase domains 1 and 2.

Figure 6.3b Mutations in FGF receptors



The positions of the mutations reported in FGFR1, FGFR2 and FGFR3 for the Pfeiffer syndrome (yellow), Apert syndrome (red), Crouzon syndrome (green), Jackson-Weiss syndrome (orange), Achondroplasia (white), Thanatophoric dysplasia (blue) and Hypochondroplasia (grey) is indicated.

transduction (Johnson and Williams 1993).

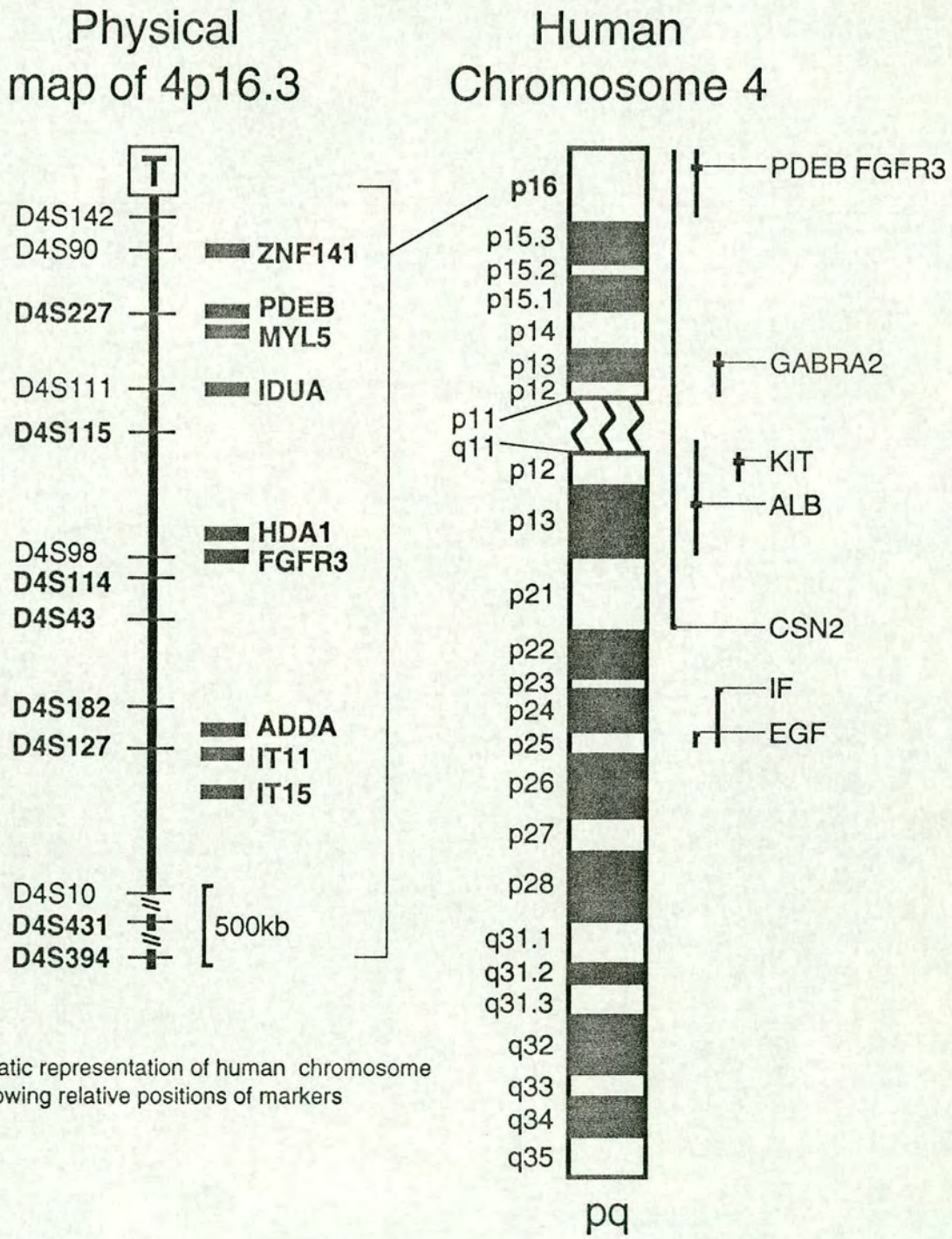
Individual FGFR genes have different patterns of expression in embryonic tissues during development (Peters et al. 1992). Even though FGFR1 and FGFR2 mRNAs are co-expressed in prebone and precartilage structures during craniofacial bone formation, FGFR3 is expressed in the cartilage growth plates of long bones during endochondral ossification of limb development (Peters et al. 1993). FGFR1 is expressed throughout the entire mesenchyme, whereas FGFR2 is confined to the ectodermal layer. FGFR4 expression is restricted to epithelium of the embryonic gut and to specific groups of skeletal muscles (Stark et al. 1991).

FGFR3 is a 125 kilo-Dalton protein and the coding region (2.5kb) has been completely sequenced (Keegan et al. 1991). Unlike the other FGFRs, the highest level of FGFR3 expression has been found in the cartilage rudiments of the developing bone. Their preferential expression during endochondrial ossification suggests that they perform differential functions during the later stages of bone development and limb formation (Peters et al. 1993). There are now at least four known skeletal dysplasias in man which result from dominant mutations in FGFR3, namely achondroplasia, thanatophoric dysplasia type I and II, and hypochondroplasia (Shiang et al. 1994, Tavormina et al. 1995, Bellus et al. 1995). The clinical spectrum of these disorders ranges from mildly affected to inevitable neonatal lethality and explains functional effects of mutations in different domains of the FGFR3 receptors. Fig 6.3b illustrates the sites of mutations in FGFR3 proteins causing these skeletal dysplasias.

6.1.4.1 *Mapping and identification of the mutation in human achondroplasia.*

Linkage analysis of human ACH with polymorphic DNA markers in two separate studies located the gene causing ACH at the telomeric region of the short arm of chromosome 4 (p16.3) (Le Merrer et al. 1994, Velinov et al. 1994). The physical map of human 4p16.3 is shown in Fig 6.4a. ACH was shown to be tightly linked to the two microsatellite flanking markers D4S412 and IDUA. Later, with multipoint linkage analysis using 10 short tandem repeat polymorphic markers, the gene causing ACH was placed in a 2.5Mb region between D4S43 and the telomere ie. telomeric to D4S412 (Francomano et al. 1994). Because ACH carriers have reduced cartilage

Fig. 6.4 a b



Diagrammatic representation of human chromosome 4p16.3 showing relative positions of markers

growth in their long bones, but normal cartilage in the remaining epiphyses, an implication of a possible abnormality of growth factor or its chondrocyte receptors as a candidate gene for ACH was suggested.

The gene encoding FGFR3 had previously been localised in this region, and was initially regarded as a potential candidate for Huntington's disease (Thompson et al. 1991). With this information, Shiang et al. (1994) examined the transcripts from this locus for alterations in ACH heterozygotes and homozygotes. DNA sequence analysis revealed a G to A transition at nucleotide 1138 of the published cDNA sequence (Keegan et al. 1991, Thompson et al. 1991), a G to C transversion at the same nucleotide position was also noted in some cases. Both mutations resulted in the substitution of the amino acid arginine for glycine at position 380 of the mature protein within the transmembrane region of the gene product. The same mutation was also detected by Rousseau et al. (1994) for ACH. Another mutation has also been found in an ACH patient, at position 375, resulting in a glycine to cysteine substitution (Superti-Furga et al. 1995).

6.1.4.2 *Thanatophoric dysplasia.*

Thanatophoric dysplasia (TD), another lethal disorder associated with skeletal dysplasia in man is characterised by micromelic shortening of limbs, macrocephaly, reduced height of vertebral bodies and shortened ribs resulting in a reduced thoracic cavity and a bell-shaped abdomen (Shah et al. 1973). The condition is classified as type I or type II, depending on the length of the femur and involvement of the skull (Tavormina et al. 1995). TD and ACH share several phenotypic features (Table 6.2). Mutations in both the extracellular and intracellular tyrosine kinase domains of FGFR3 have been detected in TD patients (reviewed in Tavormina et al. 1995). Recently, a mutation in the stop codon of FGFR3 has been demonstrated as a cause for TD type I. Three different heterozygous base substitutions in the chain termination result in the mRNA being translated a further 423bp until another stop codon is reached (Rousseau et al. 1995).

Table 6.2 Phenotype and mutations in FGFR related skeletal disorders.

	HCH	ACH	TD	PS	AS	JWS	CS
Phenotype							
Short limb dwarfism	+	+	+	-	-	-	-
Clover leaf skull	-	-	+	+	-	-	-
Under developed midface	-	+	+	+	+	+	+
Craniosynostosis	-	-	-	+	+	+	+
Foot anomalies	-	-	-	+	+	+	-
Hand anomalies	-	-	-	+	+	-	-
FGFR gene mutated							
FGFR1				P252R			
FGFR2				C342R C342Y T341P	S252W P253R	A344G	C342Y Y340H S354C S347C Y328C C342R C342S
FGFR3	N540K	G380R G375C	R248C K650E S371C				

The main clinical features defining the specific disorders are indicated by plus or minus signs. The published FGFR mutations is given with the predicted amino acid changes.
Abbreviations used: FGFR - Fibroblast Growth Factor Receptor; HCH - Hypochondroplasia; ACH - Achondroplasia; TD - Thanatophoric Dysplasia; PS - Pfeiffer Syndrome; AS - Apert Syndrome; JWS - Jackson Weiss Syndrome; CS - Cruzon Syndrome; G - Glycine; C - Cysteine; E - Glutamic Acid; S - Serine; A - Alanine; Y - Tyrosine; T - Threonine; W - Tryptophan; H - Histidine; P - Proline; R - Arginine; N - Asparagine; K - Lysine.

6.1.4.3 *Hypochondroplasia.*

Hypochondroplasia is an autosomal dominant skeletal dysplasia in man with skeletal features similar to but milder than those seen in ACH (Walker et al. 1971, Hall and Spranger 1979). McKusick et al. (1973) proposed that ACH and hypochondroplasia are allelic. Later there was genetic linkage evidence that mapped hypochondroplasia and ACH to the same locus (Le Merrer et al. 1994, Hecht et al. 1995). Hypochondroplasia has now been shown to be caused by a mutation in the nucleotide at position 1620 in the tyrosine kinase domain of FGFR3 with C to A transversion resulting in an asparagine to lysine substitution (Bellus et al. 1995).

6.1.5 Craniofacial defects in man associated with mutations in other FGFRs.

Results of clinical and molecular investigations suggest the occurrence of mutations in three members of the FGFR family as a cause of a number of malformation syndromes. FGFR3 is mainly seen associated with skeletal defects (Shiang et al. 1994). But mutations in FGFR1 and FGFR2 have recently been shown to cause several craniofacial and digital defects in man. For example, mutations in the third immunoglobulin domain of FGFR2 cause Crouzon, Jackson-Weiss and Pfeiffer syndromes (Jabs et al. 1994, Reardon et al. 1994, Lajeunie et al. 1995, Oldridge et al. 1995, Rutland et al. 1995). Interestingly, a mutation in the extracellular domain of FGFR1 is also known to cause Pfeiffer syndrome (Muenke et al. 1994). Also, two specific mutations in the linking region between the second and third immunoglobulin domain of FGFR2 cause Apert syndrome, a craniofacial defect with both mild or severe syndactyly (Wilkie et al. 1995). Mutations in human FGFR genes and malformations are shown in Fig 6.3b and Table 6.2. Some of these syndromes eg: Crouzon and Apert are allelic, and all have craniosynostosis as a common feature. Mutations in different FGFRs suggest that impaired ligand-induced receptor dimerization and signalling may be the cause of these defects (Rousseau et al. 1994). The expression of FGFR2 in developing flat bones of the skull and in the apical ectodermal ridge of the developing limb bud during early embryogenesis in mouse (Orr-Urtreger et al. 1993) may explain both the craniofacial abnormalities and the digital anomalies in these syndromes. However, the reason for different point mutations in FGFR2 causing the different

skeletal defects seen in Crouzon syndrome and Jackson-Weiss syndrome is still unclear.

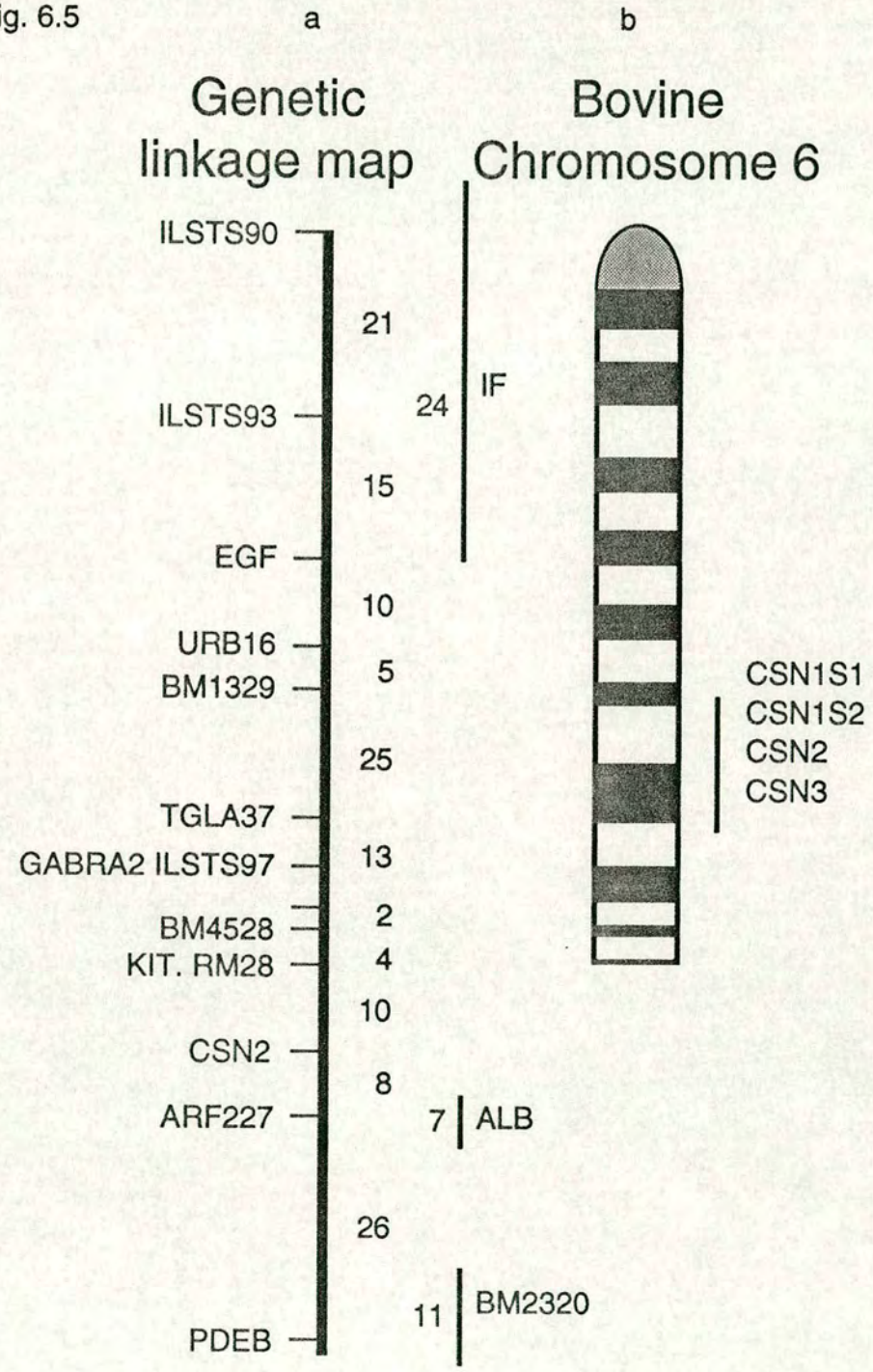
6.1.6 Synteny of human chromosome 4p to bovine chromosome 6.

The mapping of the gene responsible for ACH in man to the telomeric region of chromosome 4 and the phenotypic similarity of ACH and the Dexter dwarfism has engendered the speculation that the same gene is responsible for dwarfism in cattle. Synteny mapping in cattle has shown that the short arm of human chromosome 4 is homologous to bovine chromosome 6 (formerly bovine U15) with a high degree of conservation of coding and non-coding loci (Zhang et al. 1992) (Fig 6.5a, Fig 6.4b). The PDEB gene located in the subtelomeric region of human chromosome 4 has been mapped near the telomere on bovine chromosome 6 (Barendse et al. 1994). Similarly, other coding regions like KIT located on human chromosome 4 have also been shown to be located on bovine chromosome 6 (Fig 6.5a, 6.4b). This extensive conservation of synteny has now been confirmed by Solinas-Toldo et al. (1995) by hybridisation of human DNA to bovine chromosome in Zoo FISH studies. These experiments show that the organisation of human chromosome 4 is almost conserved in bovine chromosome 6, but with small regions present on bovine chromosomes 17 and 24. Although the genes found in human chromosome 4 are found on bovine chromosome 6, the order of the genes is not known.

6.1.7 Aim of the project.

The gene causing dwarfism being associated with a general skeletal defect, is of great interest in understanding development. Research has been conducted in man over many years to isolate the gene causing dwarfism and thus to identify a marker for the effective diagnosis of the carriers. During the first part of the project, the gene causing human achondroplasia had just been mapped to chromosome 4p16.3 and no mutations in FGFR3 had been identified. The initial aim of the project was to detect linkage between Dexter dwarfism and bovine genes that were thought to be syntenic to human chromosome 4p16.3. Because no microsatellite marker had been mapped to the telomere of bovine chromosome 6, use of human markers were investigated for linkage analysis (Huntington's disease markers). When the mutation causing ACH in man was

Fig. 6.5


a
b
Genetic linkage map
Bovine Chromosome 6
ILSTS90
21
ILSTS93
24
IF
15
EGF
10
URB16
BM1329
5
CSN1S1
CSN1S2
CSN2
CSN3
25
TGLA37
GABRA2 ILSTS97
13
2
BM4528
KIT. RM28
4
10
CSN2
8
ARF227
7
ALB
26
11
BM2320
PDEB

reported, the project changed to see if the gene and mutations responsible for dwarfism in cattle was the same as for ACH in man. The aim of the work described in this chapter then became to examine FGFR3 to see if it was the underlying cause of dwarfism in Dexter cattle.

6.2 MATERIALS AND METHODS.

6.2.1 Animals.

Families of half sib matings from six carrier bulls (short-legged) were collected from breeders for this study. Blood samples of each animal in the family were received through the Dexter Cattle Society. Each family consisted of carrier (short-legged) or non-carrier (long-legged) progeny from carrier bulls and non-carrier dams. Calves of unknown phenotype were not considered for analysis. All the animals were blood typed at the Cattle Blood Typing Service.

6.2.2 Analysis of DNA using markers on human chromosome 4p16.3.

DNA was extracted either from blood, semen or tissue samples as given in chapter 2 and was amplified using the markers located on human chromosome 4p16.3. The primers flanking the trinucleotide repeat in the HD region, HD1/3 and a 42bp repeat in the coding region of the gene HDA 1.1, and two other markers, an intragenic region within the iduronidase gene (IDUA) and D4S115 shown to be linked to the gene causing ACH in man were used for PCR amplification of bovine DNA. The DNA samples from the different bovine families were amplified with the above primers and analysed in the automated DNA fragment analyser. The procedure for designing the primers, general conditions for PCR and the method of electrophoresis have been detailed in chapter 2. The markers used, primer sequences and the reaction conditions are given in Table 6.3.

6.2.3 Linkage analysis with microsatellite markers.

Initially no polymorphic markers had been mapped to the telomeric region of bovine chromosome 6. In the latest map of the bovine genome (IBRF 14, Barendse personal communication, August 1995), some microsatellite markers are shown on

Table 6.3 Markers from human chromosome 4p16.3.

Locus	Primer sequence	PCR prog.	PCR reaction mix	Reference
HD1/3	F GAAGGCCTTCGAGTCCCTCAAGTCCTTC R GCGGTGGCGGCTGTTGCTGCTGCTGCTG	1	A	Warner et al. 1993
HD1.1	F GGCCGAGCGGCGGGACTCC R CTGTCGGTTGCGGCGGCCGC	2	A	McCombie et al. 1992
D4S115	F ATCAAGTCGAGGGACCTGGGCT R CAGACAGCAGAGTCCACGGACAG	3	B	Gusella et al. 1992
IDUA	F GGATCCCTCTGGCATCTC R TACTATCATTACCCATCC	4	C	,,

PCR reaction mix:

A. 200uM dNTP, 10mM Tris, 50mM KCl, 1.25mM $MgCl_2$ and 0.5 units Taq polymerase.

B. 1.25mM dNTPs, 67mM Tris, 16.6mM Ammonium Sulphate, 10mM BME, 10%DMSO and 0.02 unit/ul Taq polymerase.

C. 200uM dNTP, 10mM Tris, 50mM KCl, 1.5mM $MgCl_2$, 0.01%gelatin and 0.02 unit/ul Taq polymerase.

PCR programmes:

1. 4'@94°C, 35X (30s@94°C, 30s@65°C), 45s@72°C

2. 4'@94°C, 35X (30s@94°C, 30s@60°C), 45s@72°C

3. 2'@94°C, 30X (1'@94°C, 1'@60°C, 1'@72°C)

4. 2'@94°C, 30X (1'@94°C, 1'@57°C, 3'@72°C)

bovine chromosome 6. In man, FGFR3 is shown to be linked to PDEB as represented in Fig 6.4a. The microsatellite markers BM2320 which maps 11cM from PDEB, and more distant markers BM4528 and RM28, ILSTS97, and AFR227 were typed on Dexter pedigrees and tested for linkage analysis with FGFR3. The primer sequences and the reaction conditions are given in Table 6.4.

The program CRI-MAP was used for two-point linkage analysis with microsatellite markers. The affected individuals were scored as heterozygotes at the index locus while unaffected individuals was scored as homozygotes for the normal allele. The program then computed the lod score between the two loci at recombinant fractions ranging from 0.001 to 0.5.

6.2.4 Isolation of bovine specific sequence for the transmembrane region of FGFR3.

Primers were designed for the 7th, 8th, 9th, 10th and 17th exon of the FGFR3 gene based on the published human sequence in order to detect polymorphisms (Keegan et al. 1991). The primer sequences and the reaction conditions are given in Table 6.5. A schematic representation of the exons and the locations of the different primers are shown in Fig 6.6.

6.2.4.1 *Subcloning of PCR products*

PCR products were subcloned using a TA cloning system (Invitrogen). After ligating the PCR products into the TA vector, the resulting clones were transformed into competent *E.coli* cells and were plated on L agar plates with X-gal and ampicillin. Recombinant white colonies were selected and preliminary screening of the colonies was carried out by PCR amplification of an aliquot of liquid culture using the primers flanking the insert. Cultures of cells containing vector with the correct size of insert were grown and plasmid DNA prepared using the Wizard miniprep system (Promega). Restriction digestion with *Eco* RI was performed on the miniprep DNA to release the insert. The protocols for TA cloning, PCR screening of miniprep cultures, preparation of plasmid DNA, restriction endonuclease digestion and electrophoresis are given in detail in chapter 2.

Table 6.4 Microsatellite markers on bovine chromosome 6 used for linkage analysis.

Locus	Primer sequence	PCR Prog.	Mg Conc.	Reference
BM2320	F GGTTCCCAGCAGCAGTAGAG R CCCATGTCTCCCGTTACTTC	1	1.25mM	Bishop et al. 1994
BM4528	F CAGAATCCATACACATGTCAAC R AGGAACAGGTATAGGAATATTGG	1	1.25mM	"
RM028	F CTACAGTCATGGGTCTGAAAGAG R ATCTTCAGCCTGGCCTGAGAG	1	1.25mM	Barendse et al. 1994
ILSTS97	F AAGAATTCCCGCTCAAGAGC RGTCATTTCACCTCTACCTGG	2	1.25mM	Accessno: L3741
AFR227	F GACCAACAGAGTGCATGCACG RTCATTGAGCAGGAGTAGGATTG	1	1.5mM	Jorgensen et al. 1995

PCR programmes:

1. 2'@92^{0}C, 30X (30s@94^{0}C, 30s@58^{0}C, 30s@72^{0}C)

2. 2'@94^{0}C, 30X (30s@94^{0}C, 30s@55^{0}C, 30s@72^{0}C)

Table 6.5 Primer pairs in different exons of FGFR3.

Code	Exon	Primer sequence	PCR prog.	Mg conc. (mM)
C1	9	F GGGGCTTGCCCAGGGTCAGC	1	1.25
E1	10	R CAAATGGGAGCTGTCTCGGG		
B1	8	F TGGTGGAGGCTGACGAGGCG	1	1
D1	8	R TGTCGCTTGAGCGGGAAGCG		
B2	8	F AGGAGCTGGTGGAGGCTGA	1	1
D2	8	R GGAGATCTTGTGCACGGTGG		
C1	9	F GGGGCTTGCCCAGGGTCAGC	2	1.25
F1	17	R CGTCACCCCCGAGCGCCTGC		
A1	7	F GCGTGGCTGGTGGTGCTGCC	1	1.25
E1	10	R CAAATGGGAGCTGTCTCGGG		
A1	7	F GCGTGGCTGGTGGTGCTGCC	2	1.25
F1	17	R CGTCACCCCCGAGCGCCTGC		
B2	8	F AGGAGCTGGTGGAGGCTGA	2	1.25
F1	17	R CGTCACCCCCGAGCGCCTGC		
A1	7	F GCGTGGCTGGTGGTGCTGCC	1	1
D2	8	R GGAGATCTTGTGCACGGTGG		
B2	8	F AGGAGCTGGTGGAGGCTGA	1	1.25
E1	10	R CAAATGGGAGCTGTCTCGGG		
B3	8	F CCGAGGAGGAGCTGGTGGAGGCTGACGAGG	3	1.25
D3	8	R GAGCGGGAAGCGGGAGATCTTGTGCACGG		
B4	8	F GCAGTGTGTTCGCGGGTGTCCTCAG	4	1.25
D4	8	R CGAGCCCAGGCCCTTCTTGG		

PCR programmes:

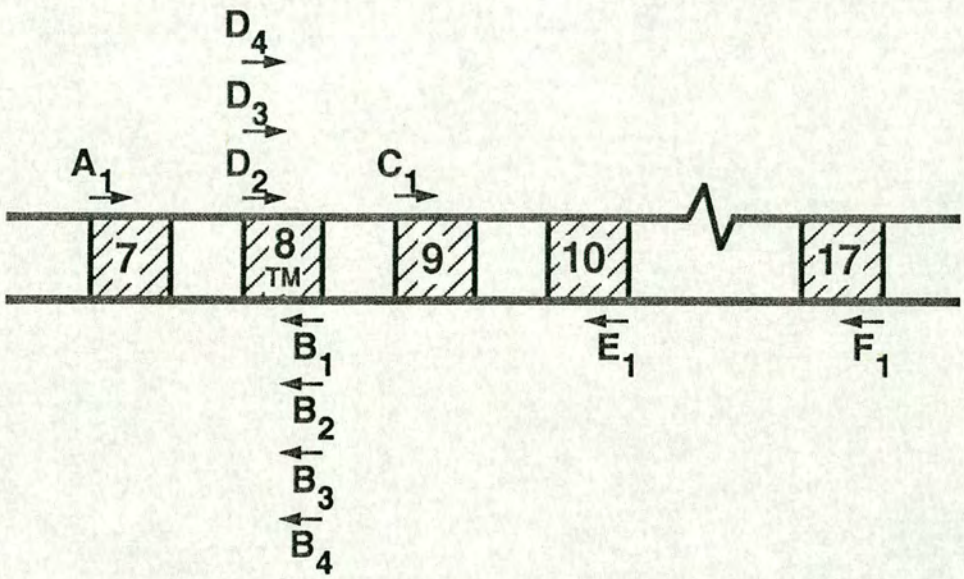
1. 2'@94^{0}C, 30X (1'@94^{0}C, 1'@60^{0}C, 1'@72^{0}C).

2. 2'@94^{0}C, 30X (1'@94^{0}C, 1'@60^{0}C, 1'@72^{0}C), 5'@72^{0}C.

3. 3'@94^{0}C, (1'@94^{0}C, 1'@67^{0}C, 1'@72^{0}C), (1'@94^{0}C, 1'@66^{0}C, 1'@72^{0}C), (1'@94^{0}C, 1'@65^{0}C, 1'@72^{0}C), (1'@94^{0}C, 1'@64^{0}C, 1'@72^{0}C), (1'@94^{0}C, 1'@63^{0}C, 1'@72^{0}C), (1'@94^{0}C, 1'@62^{0}C, 1'@72^{0}C), (1'@94^{0}C, 1'@61^{0}C, 1'@72^{0}C), 20X (1'@94^{0}C, 1'@60^{0}C, 1'@72^{0}C), 5'@72^{0}C.

4. 2'@94^{0}C, 30X (1'@94^{0}C, 1'@65^{0}C, 1'@72^{0}C).

Figure 6.6 Schematic representation of the location of primers in different exons of human FGFR3



Each primer is represented by a letter and the direction of primers in each exon is indicated by arrows. Primers redesigned for the same exon are indicated sequentially.

6.2.4.2 *Hybridisation and Southern blotting.*

The PCR product, amplified with primers flanking the TM region of FGFR3, was run on an agarose gel and the band was excised to separate the product from the primer band. The gel slice containing the PCR product was used as a probe for hybridisation. The methods followed for blotting, megaprime labelling of the probe and hybridisation are described in chapter 2.

6.2.4.3 *Screening of the lambda gt11 library.*

A bovine phage lambda gt11 cDNA library made from stimulated lymphocytes was available in the laboratory (Fraser et al. 1994). The library was screened for the presence of a bovine FGFR3. The titre of the phage was first determined as 5 X 10^7 by plaque assay as described in chapter 2. An 1/300 dilution of this phage was used to obtain 1.5 X 10^5 plaques/plate. The screening of the library was done in large NZY plates (20cm X 20cm) to screen approximately one million plaques. The phage particles were blotted on to nitrocellulose filters in duplicate. The DNA of the phage particles was denatured, neutralised, fixed and hybridised with the probe made from the transmembrane region of FGFR3. The plaques which hybridised were picked for secondary screening.

In secondary screening, the plaques were identified on the plates with the information from the autoradiographs by proper alignment of the filters. The plaques were then marked with a needle, withdrawn into a pastette and transfered to an Eppendorf with 500ul of SM buffer to elute particles. The titre of the phage was determined and screening was repeated to obtain individual plaques. The phage DNA prepared was then used for PCR amplification and sequencing.

6.2.4.4 *Sequencing.*

The plasmid/phage DNA was sequenced by different procedures. The DNA was sequenced by the sequenase method (Amersham). If the amount of DNA was less than 3-4ug, cycle sequencing was done using vent DNA polymerase (New England Biolabs). The reactions were loaded onto 6% denaturing polyarylamide gels and electrophoresed

for 2-3 hrs. The methods of cycle sequencing, sequenase reaction and denaturing polyacrylamide gel electrophoresis are explained in detail in chapter 2.

6.2.4.5 *Touchdown PCR.*

Touchdown PCR was carried out as described by Don et al. (1991) using long primers in the 8th exon of FGFR3 flanking the area of mutation causing ACH. This method improves the specificity of the PCR and reduces the formation of spurious products. The primers used in touchdown PCR are presented in Table 6.5. The PCR conditions used were as follows. An initial denaturing step at 94^0C for 1min. was followed by an annealing reaction at 67^0C and an extension reaction at 72^0C for 1min. At each subsequent cycle the annealing temperature was reduced by 1^0C until it reached 60^0C, after which 20 cycles were performed at the same annealing temprature. The final extension step was at 72^0C for 5min. The products of touchdown PCR were analysed on a gel, and hybridised with the probe made from the TM region of FGFR3. The products giving a positive signal were subcloned and sequenced.

6.3 RESULTS.

6.3.1 Linkage analysis using human markers.

After the initial localisation of human ACH to 4p16.3 (Le Merrer et al. 1994, Velinov et al. 1994) the syntenic bovine chromosome 6 was studied for potential markers to be used in the Dexter dwarf pedigrees. However, on the bovine map available at that time (IBRF 8), the telomeric region of chromosome 6 had no microsatellite markers. Later, PDEB was added to the international bovine genome map by Barendse (IBRF9) which was mapped telomeric in chromosome 6 identifying the approximate location of FGFR3. This marker was not found to be informative in any of the Dexter pedigrees that were investigated (Barendse pers. comm.). Therefore, linkage analysis on the Dexter pedigrees using polymorphic conserved markers present in human chromosome 4p16.3 was pursued to obtain a clue as to whether the gene causing dwarfism in Dexter cattle is also located in the same syntenic region as ACH in man. Two markers, IDUA and D4S115, known to be linked to the gene causing ACH in man were used. In addition, human primers were designed to amplify the

polymorphic trinucleotide repeat in the Huntington's gene HD 1/3 and a 42bp repeat in the coding region of the neighbouring gene HDA 1.1.

Even though human DNA could be amplified using the microsatellite marker D4S115, no PCR product was obtained for bovine DNA. Persumably, the primer sequences for this locus are not conserved in cattle. Bovine DNA could be amplified with the marker IDUA, but the PCR products were not polymorphic. The marker HD 1/3 was found to be monomorphic in three Dexter families but was diallelic in the other families. Analysis of these markers was difficult because the product size was only 70-80bp and interference from the primer front when analysed on the automated sequencer made interpretation difficult. The samples were therefore electrophoresed on agarose gels and melted agarose fractions from the specific bands were loaded onto the automated sequencer. However, only one family was found to be informative with HD 1/3 and HDA 1.1 markers. The product after one round of PCR with bovine DNA was too faint to be seen clearly on agarose gels for HDA 1.1. A second round of PCR always yielded a stronger product but also gave non specific products. The HDA 1.1 marker was also found to be diallelic in the Dexter pedigrees and the alleles were only 10bp different in size indicating a change in the size of the repeat. However, none of the six Dexter pedigrees was informative because the carrier bulls were homozygous for this marker.

FGFR3 has been mapped to human chromosome 4p16.3 and dwarfism being a defect in growth and development suggests a possible role of FGFR3 (Le Merrer et al 1994, Velinov et al. 1994). An attempt was therefore made to detect a polymorphism in bovine FGFR3 for linkage analysis. However, because the bovine sequence for FGFR3 was not available, primers were designed based on the published human sequences for FGFR3 (Keegan et al. 1991) for the 7th, 8th, 9th, 10th and 17th exons to amplify the intron sequences in order to detect polymorphism by SSCP or DGGE. Fig 6.6 indicates the location of primers for amplification of specific regions. The primers designed were coded for easy identification: forward primers were designed for the 7th (A1) and 9th (C1) exons, and reverse primers for the 8th (D1), 10th (E1) and 17th (F1) exons. Areas of maximum homology between the FGFR3 sequences of mouse and man was used in designing the primers on the assumption that they would

be conserved in cattle. The sequences of the primers designed in the different exons are given in Table 6.5. No PCR product was obtained for any of the bovine DNA samples with the primer pairs A1/F1 and C1/F1. Products of the predicted size were seen for the control human DNA sample with all these primers. Even though bovine DNA could be amplified with the other primers, the product was always bigger than that from human DNA, and also gave other non-specific products when compared with the human product. As the primers were not species-specific, sequence mismatches could cause difficulties in amplification.

6.3.2 Linkage analysis with microsatellite markers.

A new microsatellite marker BM2320 was later mapped near the putative bovine FGFR3 locus on chromosome 6 and was used for linkage analysis in available Dexter pedigrees to exclude or include FGFR3 as a candidate gene for dwarfism. This marker mapped 11cM proximal to bovine PDEB and hence was likely to be linked to bovine FGFR3, which also maps proximal to PDEB in man (Fig 6.4a, 6.5a). Four of the six families were informative for this marker. But in three of the families, this marker did not follow mendelian segregation of alleles, indicating an error in the pedigree. Therefore the priority was to verify the pedigrees before more work was carried out.

Additional markers have been added to the map of chromosome 6 in the latest published bovine genome map. The position of the markers in chromosome 6 are indicated in Fig 6.5a. Some of the markers were known to be closely linked to each other. If the pedigree structure was correct, close linkage would be demonstrable between these markers. Therefore, the linked markers ILSTS97, BM4528, RM28 and AFR227 were typed in the Dexter families to check the family structure, and if correct an effort was to be made to findout whether these markers were linked to the mutation causing dwarfism in Dexter cattle. Some families were informative for some markers and not for others. Errors in parentage were detected in five of the six families, with 3-4 calf-dam pairs incorrectly identified. In one family, out of 14 calf-dam pairs, 10 proved to be incorrect!. Two point linkage analysis was done between the markers known to be linked and also between each marker and the gene causing dwarfism. The program CRI-MAP was used for analysis on families showing consistent inheritance.

Even those markers which were known to be closely linked on the International bovine genome map were found to be unlinked when analysed in the Dexter pedigrees. For example, the markers RM28 and BM4528 are known to be linked (4cM) but showed a lod score of -2.35 at recombination fraction of 0.05 when tested on Dexter pedigrees. As a lod score of -2 is taken as significant exclusion, the results are not consistent with the expected mapping data. Moreover, blood typing results also showed parentage errors in some families. Therefore the pedigrees were used not further.

6.3.3 Isolation of bovine FGFR3 transmembrane sequence.

As mentioned in section 6.1.4.1, three mutations in the transmembrane region of FGFR3 have now been shown to be responsible for cases of ACH in man (Shiang et al. 1994, Superti-Furga et al. 1995). To determine whether any of these mutations in FGFR3 was the cause of Dexter dwarfism the TM region of FGFR3 was isolated from cattle and sequenced. To amplify the TM region from bovine, DNA primer pairs used were those designed in the 8th exon of FGFR3 (B1/D2) based on the area flanking the mutation in man (Table 6.5). However, no PCR product was obtained with bovine DNA using these primers. Then the primers by those of Shiang et al. (1994) to detect the mutation in the TM region of human FGFR3 (B2/D2) were used and their sequences and the reaction conditions are presented in Table 6.5. Faint products could be amplified from the bovine DNA with sizes similar to those of the human products. The PCR products from normal bovine and human DNA were subcloned into the TA vector and four clones each from the bovine and human clonings were sequenced. All the human and bovine clones had exactly the same sequence as the published human sequence for FGFR3, indicating either a complete homology between human and bovine sequences which is unlikely, or contamination of human DNA in the DNA from Dexters.

Therefore, to differentiate between human and bovine products, a bovine cDNA library (stimulated lymphocyte) was screened by PCR using the primers from the 7th and 8th exons (A1/D2). Any human DNA contamination would be revealed as a full length genomic DNA product (as these primers flank an intron), while a shorter product would be expected for the bovine cDNA. A human genomic DNA sample was used

as a positive control. An ~600bp product was seen for the human genomic DNA and a smaller faint product of 200bp for the cDNA library. Second round PCR amplification of the 200bp product with nested primers (primers in the 8th exon (B1/D2) to amplify the transmembrane region) resulted in a strong clean product. The gel was blotted and hybridized with the probe made from the human sequence of TM region of FGFR3. The product from the cDNA library showed a positive signal indicating the presence of FGFR3 in the library (Fig 6.7a).

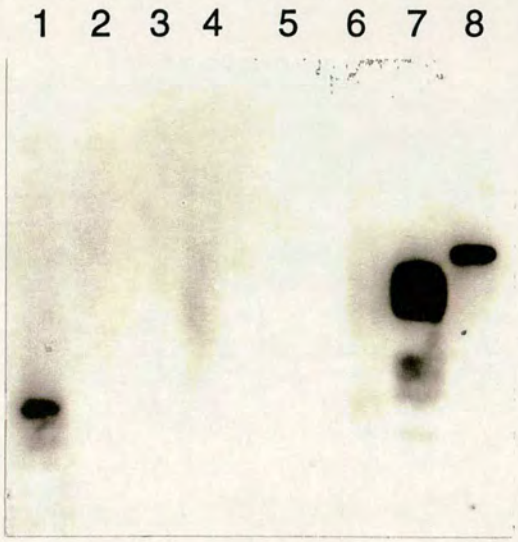
6.3.3.1 *Screening of the Phage library.*

The cDNA library was screened to isolate the bovine TM region of FGFR3 for sequencing. The filters from the primary screening were hybridised with the TM region of FGFR3 and four plaques showed a positive signal, all of which showed a positive signal on secondary screening. This finding suggests that either lambda gt11 or the bacterial strain (Y1088) had some FGFR3 homologous sequence or more unlikely that all plaques are true positives. The primer sequences were compared for homology with phage lambda and the lac operon of the plasmid in Y1088, using sequences obtained from GENBANK. The primers showed 83 and 78% homology respectively with lambda phage and lac operon, indicating that the product obtained from the phage library was likely to be either from gt11 or Y1088.

6.3.3.2 *Touchdown PCR.*

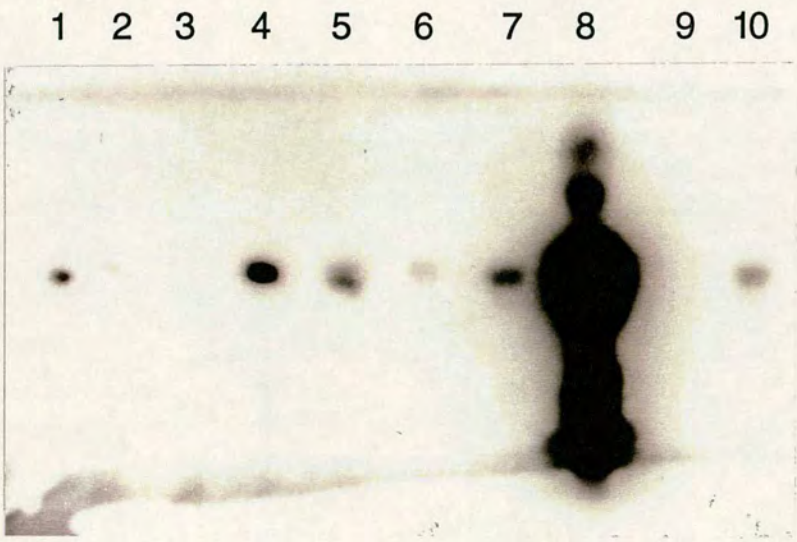
The attempts to isolate the bovine sequence for FGFR3 from genomic DNA, cDNA and phage DNA library failed because of the sequence homology between the primers and the vector. Longer PCR primers (29 and 30 oligonucleotides, B3/D3) were therefore designed in the TM region flanking the area of mutations based on the human sequence. The primers were designed in such a way that there was less homology with lambda or *E.coli* sequences. Bovine genomic DNA, bovine cDNA freshly prepared from activated macrophage and B cell lines, and phage DNA prepared from the lambda gt11 library were amplified under various PCR reaction conditions. Strong PCR products were seen for all the reactions, including the negative control. The gel was blotted and on hybridisation with TM region of FGFR3, all the lanes showed a positive

Figure 6.7a Autoradiograph of the Southern blot showing presence of FGFR3 in phage library



PCR products from the phage library probed with ^{32}PdCTP labelled segment of TM region of FGFR3. Lane 1 - phage DNA, Lane 7 - marker, Lane 8 - genomic DNA.

Figure 6.7b Autoradiograph of the Southern blot of the gel after Touchdown PCR



Products after Touchdown PCR probed with ^{32}PdCTP labelled segment of TM region of FGFR3. Lane 1- cDNA from B cell line, Lane 4 - cDNA from activated macrophage, Lane 5 - genomic DNA from "bull-dog" calf, Lane 6 - genomic DNA from a non-carrier, Lane 7 - genomic DNA from a carrier, Lane 8 - marker, Lane 10 - phage DNA

signal.

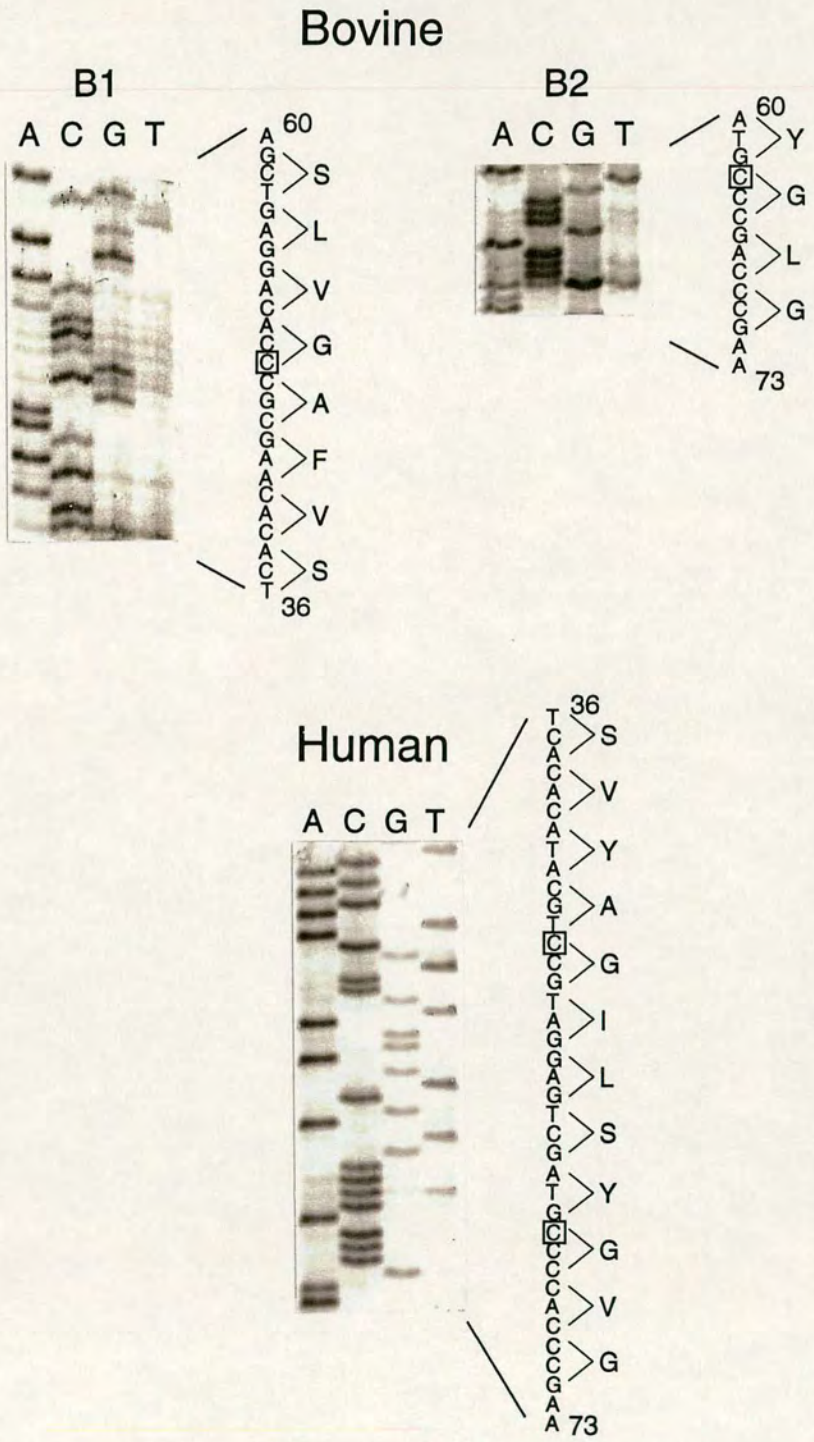
The technique of touchdown PCR (Don et al. 1991) was therefore used with the newly designed primers, whose sequences and PCR conditions are shown in Table 6.5. Because the PCR cycling programme started at a very high annealing temperature, only very specific products were initially amplified. A PCR was set up using bovine genomic DNA, phage library DNA, cDNA (activated macrophage and B cell line) and a positive and negative control. All the samples except the negative control showed the correct size PCR product. When blotted and hybridised the gel showed a strong signal for the cDNA and human genomic DNA products but weak signals for all other samples (Fig 6.7b).

The strong PCR products from the human genomic DNA and the bovine cDNA were subcloned into the TA vector. A total of 80 clones from the different transformations were chosen and screened by colony PCR as described by Hawker and Billadello (1993), of which 36 clones had the correct size insert.

6.3.4 Comparison of TM region of FGFR3 in bovine, human, mouse and chicken.

The 182bp PCR products from the human and bovine cDNAs (activated macrophage and B cell line) were subcloned and sequenced with the SP6 and T7 vector primers. Three clones from human and ten clones from the bovine cDNAs (5 clones from each) were analysed. Each of the human clones agreed with the published sequence of FGFR3. Sequence of nucleotides from 36-73bp in the TM region of human FGFR3 is shown in Fig 6.8. The nucleotide and the amino acid sequences of the bovine cDNA in comparison with the sequences from human, mouse and chicken are shown in Figs 6.9a and 6.9b. The bovine nucleotide and the aminoacid sequence of the TM region of FGFR3 clearly differ from those of human, mouse and chicken. The bovine DNA sequence showed 89% homology with the human, 84.5% homology with mouse and 63.3% homology with chicken mRNA for FGFR3. The bovine sequence contained an equivalent open reading frame to the human FGFR3 gene. In the 182bp sequence, 20bp were different in comparison with the human sequence, encoding 7 aminoacid substitutions. This sequence for the TM region of bovine FGFR3 was compared with other FGFRs (FGFR1, FGFR2, FGFR4) to confirm the identity of

Figure 6.8 Autoradiograph showing comparison of sequences in the TM region of bovine and human FGFR3



Antisense strand showing sequence comparison in the TM region of bovine and human FGFR3 is shown from nucleotides 36 to 73. The nucleotide "C" at positions 48 and 63 mutated in the human causing ACH is conserved in Dexter "bull-dog" calf. The sequence difference between the bovine and human FGFR3 is indicated.

Fig 6.9a Comparison of nucleotide sequences of TM region of FGFR3 of bovine, human, mouse and chicken

```
             1                                                                                                  100

Bovine     ---------- ----g----g g----c---- -t-------- --t---g--t g--------- -----c---- ----c----- ---------- cc--------
Human      ---------- ----g----g g----c---- -g-------- ---t------ a--------- ---------- ---------g ---------- -------g--
Mouse      -t-------- ----a----a a----t---- -t-----c-- ---------- g--------- ---------t ---------- ---------- -------a--
Chicken    -a-ca----- ----a----a atg--t--tt -g---tca-- ------t--- a-t------- -t--cact-- ---ag----- ---------- --c---t-at
Consensus CCGAGGAGGA GCTG-TGGA- -CTGA-GAGG C-GGCAGTGT GTACGCAGGC -TCCTCAGCT ACGGGGTGGG CTTCTTCCTC TTCATCCTGG TGGTGGCCGC

          101                                                                                               182
Bovine     c--t-c---- -a-------a -g-------- ---------- --------g- --g------- ----a----- ---------- --
Human      -----c---- ---------c ---------- ------a--- --------c- ---------- ----a----- ---------- --
Mouse      -----ta--- ---------c ----t----- a--------- t-------g- ---------- ----g----t ---------- -t
Chicken    -----tta-- ---a-ga--a aa-tg--aaa ---a----c- a--aa-a-ca ----t--a-- g--ag----- aaa--t--a- --
Consensus TGTGA-GCTC TGCCGCCTG- GCAGCCCCCC CAAGAAGGGC CTGGGCTC-C CCACCGTGCA CAAG-TCTCC CGCTTCCCGC TC
```


Fig 6.9b Comparison of aminoacid sequences of TM region of FGFR3 in bovine, human, mouse and chicken.

```
Bovine       ----v----- ---f--v--- -l--l----a ----t-y--- ---------- a---i-----
Human        ----v----- ------i--- ---------- ----t----- ---------- ----i-----
Mouse        ----m-t--- ------v--- --v------- ----i----- ---------- ----v-----
Chicken      a---m-m-ds ------i--- -t-lv----- lvi-ii--mk m-n--amntt --q-v-k---
Consensus    EEEL-EADEA GSVYAG-LSY GVGFFLFILV VAAV-LCRLR SPPKKGLGSP TVHK-SRFPL
```

The nucleotide and amino acid sequences of TM region of FGFR3 in bovine is shown in comparison with that of human (Keegan et al. 1991), mouse (Katoh et al.1993) and chicken (Pasquale 1990). Same nucleotides or aminoacids are indicated by dashed lines.

the sequenced fragments (Fig 6.10a and 6.10b). No significant homology was obtained between the bovine TM region of FGFR3 and other FGFRs.

6.3.5 Sequencing of TM region of FGFR3 in "bull-dog" calf.

Once the TM region of FGFR3 from bovine cDNA was sequenced, bovine specific primers (B4/D4) were designed to amplify the segment of DNA flanking the site of mutation causing ACH in man. The nucleotide G at position 63 if substituted by either A or C resulted in ACH in man. Primers designed were internal to those used in touchdown PCR, but flanking the site of mutation. The sequence of the primers and the reaction conditions are given in Table 6.5. A strong PCR product was obtained for all the DNAs and the products were subcloned and sequenced. A total of 13 clones from carrier, non-carrier and "bull-dog" calf DNA were analysed, and no sequence difference was seen between any of them. The sequence of nucleotides (in the antisense strand) from 36 to 73bp in the "bull-dog" calf is shown in Fig 6.8. The nucleotide at position 48 and 63 was C in both non-carrier Dexter cattle and the "bull-dog" calf. This clearly indicates that the mutations in the TM region of FGFR3 responsible for ACH in man is not the cause of Dexter dwarfism.

6.4 DISCUSSION.

6.4.1 Linkage analysis.

At the time the project started, it was just then published that the gene causing ACH in man had been mapped to chromosome 4p16.3. Comparative mapping information showed that bovine chromosome 6 is syntenic with human chromosome 4p16.3. Because few polymorphic markers for linkage analysis were present on bovine chromosome 6, markers on human chromosome 4p16.3 were selected and tried on bovine DNA.

Multipoint linkage analysis in man placed the ACH locus within an estimated genetic distance of 4cM and distal to the Huntington's disease locus. So few polymorphic markers in man linked to Huntington's disease, HDA 1.1 and HD 1/3 (McCombie et al. 1992, Warner et al. 1993) and two other markers known to be linked to ACH, IDUA and D4S115 (Le Merrer et al. 1994, Velinov et al. 1994) were analysed

Fig 6.10a Comparison of nucleotide sequence of the TM region of human FGFR1, FGFR2, FGFR3 and FGFR4 with bovine FGFR3

```
            1                                                                                                   100
HumanFgfr1  aa-ccct--- aga-ag-cc- --a-tgat-a -ctcgc-cc- ----ct---g ---a------ -tt-c-ca-- ggc------- a--t--tgca ------g-t-
HumanFgfr2  a-t-cct--- aga-a.aa-- -agattaca- -ttc-c-a-a c---ct---g --agc---t- --t-c--a-- gg-----t-a a--g--tgta ------taa-
BovineFgfr3 c--------- gct-gt-g-- --t-a--a-- -t-g-agt-- --t-g---gt g--c---g-- -----c---- ----c----- ---at-c--g cc----c---
HumanFgfr3  c--------- gct-gt-g-- --t-a--a-- -g-g-agt-- ---tg-a-gc ---c---g-- -----g---- ---------g ---at-c--g ------c---
HumanFgfr4  ggag---ccc cacatg-acc --a-ca-c-c -c-agg-cag ---ta----c ---a--c-g- ---c-tc--- --c-c-ggc- --gg-tg--c -cc--ct-ct
Consensus   -CGAGGAGGA ---G--G-AG GC-G-CG-GG C-G-C-C-GT GTAC-CGGA- ATC-TCATCT ACGGGATGGG CTTCTTCCTC TTC-CC-TG- TGGTGG-CGC

            101                                                                                                 182
HumanFgfr1  -----tcg-- ---aaga--- ----tggta- -------agt ---t--ca-a g--agatg-- ---------- --gg--aaga g-atc-ct
HumanFgfr2  a----tc--g -g---aa--- ---a-a-ga- -------cca ---t--a--a g--agc-g-- ---------- --ga--aaac gtatc-c-
BovineFgfr3 c--t-c---- -----cc--- g--------- -------... -g---gg--t -g---g--.. .--------- a--------- ----g---
HumanFgfr3  t--g-c---- -g---cc--c gc-------- ------a... -g---gg--t -----a--.. .--------- a--------- ----g---
HumanFgfr4  --c-gg---g --t--agg-c --gcg-t--a -ggccg-... c---c-c--c -g---g--a- ------g--- ---------- ----t--g
Consensus   GGTCA-GCTC TACCG--TGA AGAGCCCCCC CAAGAAG--- GACCTC-GC- CCCCC-CCGC TGTGCACAAG CTCTCCCGCT TCCC-CTC
```


Fig 6.10b Comparison of aminoacid sequence of the TM region of human FGFR1, FGFR2, FGFR3 and FGFR4 with bovine FGFR3

```
Human Fgfr1  al-erp-vmt spl-le-ii- ct-a--isc- -gs-iv-kmk -gt--s--hs qm-----aks ip
Human Fgfr2  apgrekeit- spd-le-ai- ci-v--iac- -vt-i-c-mk ntt--p--ss q------tkr ip
Bovine Fgfr3 ---------- gs-fagvls- gl--l-fila ----t---lr --p---..-g s-----i--- --
Human Fgfr3  ---------- gs--ag-ls- gv----filv ----t-c-lr --p---..-g s-t---i--- --
Human Fgfr4  edptwt-aap ear-td-il- as-slalavl lllag---gq alhgrhp.rp pat-q----- --
Consensus    EEELVEADEA --VY--I--Y --GFFL---M VAAV-LYR-- SP-KKGDFL- -PAVHKLSRF PL
```

The nucleotide and amino acid sequences of TM region of bovine FGFR3 is shown in comparison with that of human FGFR1, FGFR2 (Dionne et al. 1990), FGFR3 (Keegan et al. 1991) and FGFR4 (Partanen et al. 1991). Same nucleotides and amino acids are indicated by dashed lines. Gaps introduced into the sequence is indicated by dotted lines.

on Dexter DNA samples. The primers in the first exon of the gene HDA 1.1 flank a 42bp sequence that have been found to be tandemly repeated two or four times (McCombie et al 1992). Cattle appear to have a smaller repeat size. The Dexter DNA samples showed polymorphism with this marker but was found uninformative in the Dexter pedigrees because all the carrier bulls were homozygous. In cattle, the marker HD 1/3 gave a product ranging 70 - 80bp. The product size for this marker in man ranged between 80bp (11 CAG repeats) to 143bp (32 repeats) in normal human chromosomes and 173bp (42 repeats) or larger for patients with Huntington's disease (Warner et al. 1993). This shows that the trinucleotide repeat in cattle is not as variable as in man. This is also the case in the equivalent Huntington's gene in the mouse (Lin et al. 1995). A third gene marker IDUA was also amplified but was not polymorphic which indicated the conservation of genes between man and cattle. The marker D4S115, not surprisingly, could not be amplified on bovine DNA as this microsatellite marker was in a non-coding region which is unlikely to be conserved between man and cattle.

PDEB is the only gene mapped to the telomeric region of chromosome 6 in cattle. Unfortunately, the PDEB marker was found to be homozygous in all of the Dexter pedigrees. Velinov et al. (1994) suggested that FGFR3 was a potential candidate for ACH by virtue of its pattern of expression. Earlier, Johnson and Williams (1993) had hinted its role in growth and development. As FGFR3 is known to be expressed exclusively in resting cartilage during endochondral ossification (Peters et al. 1993), it was assumed that linkage analysis with this marker might indicate its role in Dexter dwarfism. Primers designed from the published human sequence could not be successfully used on bovine genomic DNA for various reasons. Sequence mismatches, especially at the 3' end in primer sequence between the two species might be one of the factors hindering amplification. Between sheep and cattle, sequences are more conserved and PCR primers designed from one species can be used for the other (Moore et al. 1991) whereas fewer primers from man were expected to work in cattle.

A strong PCR product is generated in the negative control of the same size as in the genomic and cDNA controls when amplified with primers within the exons. This indicated human or bovine genomic DNA contamination in the reaction. The primers

on the 7th and 8th exons flanking an intron, when used on cDNAs always generated a similar size product with human and bovine DNA but a bigger product for bovine DNA. This suggested the presence of a bigger intron.

When microsatellite markers became available on bovine chromosome 6, Dexter families were typed with these markers for use in linkage analysis to exclude/include FGFR3 as a candidate for dwarfism in Dexter cattle. Even though the markers were informative in the families, they could not prove linkage because of errors in the pedigrees. The results of linkage analysis showing a negative lod score for closely linked markers, proved that these families are not useful for further linkage analysis. New families will be identified to continue this approach. Apart from the paternity errors, there may have been mistakes in the diagnosis of carrier/non-carrier status of the defect, which might have skewed the results in linkage analysis.

6.4.2 Bovine sequence of TM region of FGFR3.

Sequencing of the transmembrane region of FGFR3 in bovine DNA using human specific primers was at first unsuccessful due to human DNA contamination and primer mismatches. The primer initially designed based on the human sequence, had a mismatch with the mouse sequence at the 3' end and failed to give any PCR product on amplification. It is possible that this mismatch was also present in the bovine sequence and this might have accounted for the primers failing to give any PCR products. But redesigning the primer by avoiding the mismatch at the 3' end, the PCR amplification was improved. It was therefore clear that the conservation of homology between different species at the 3' end of primers needs to be considered in designing the oligonucleotides for efficient PCR amplification of unknown sequences.

PCR amplification with human-specific primers of bovine genomic DNA, cDNA derived from bovine RNA, and bovine cDNA library always gave a human-specific product, as revealed by sequencing. The primers designed from human sequence seemed to anneal more readily to contaminating homologous human DNA in the reaction than to the bovine DNA suggesting variation in the bovine sequence. The contamination could have occured either during the preparation of bovine DNA or in the reagents or by air-borne transfer. Normally, within the laboratories these problems

do not occur as bovine specific primers are used. This problem of human DNA contamination was encountered only when using human specific primers on any DNA from a different species. Because of the problem of human DNA contamination, it was necessary to determine the bovine sequence for the TM region of FGFR3 by a more difficult procedure.

FGFR3 was shown to be present in the bovine lambda phage library by Southern blotting and hybridisation of the PCR products. Because the full length of the cDNA probe for FGFR3 was not available, the library was screened with a human probe prepared from the PCR product amplified across the TM region. However, this 182bp probe proved to be too small. Moreover, the primers used for amplification had homology with sequences in both the vector and the bacteria used for propagation, leading to hybridisation of these regions instead of the bovine sequence for FGFR3. To overcome this a larger probe would have been better for screening the library.

The use of touchdown PCR (Don et al. 1991) on bovine DNA, using longer PCR primers which were human-specific, helped to amplify a bovine-specific product. Longer primers gave a greater chance of priming in the correct bovine sequence by allowing some base pair mismatches. Touchdown PCR is a very sensitive tool which helps to overcome the problem of contamination and non-specific primer annealing.

Sequencing the 182bp PCR product after touchdown PCR revealed a highly GC rich sequence which was comparable to the sequence of the TM region of human FGFR3. The sequence contains runs of C's, G's and GC's which were always stacked together and could not be resolved when the sequencing gel was electrophoresed at standard temperature of 40- 50^{0}C, and at a higher temperature (65^{0}C) the stacked bands were separated.

The bovine sequence for the TM region of FGFR3 in comparison with human, mouse and chicken sequences showed differences of 20, 28 and 72bp respectively. At the amino acid level, the bovine sequence differed from human, mouse and chicken sequence by 7, 9 and 27 residues respectively. In all the cases, these substitutions were for uncharged amino acids and were replaced by other amino acids with similar properties (classification of aminoacids Lehninger 1982). It is interesting to note that the glycine at positions 375 and 380 in the human, mouse and chicken sequences was

also conserved in cattle (Johnson and Williams 1993, Ornitz and Leder 1992, Pasquale 1990).

In ACH patients, the glycine residue at positions 380 and 375 are substituted by arginine and cysteine respectively by alterations in the nucleotides at position 1138 from G to an A or C, and at 1122 from G to T (Shiang et al. 1994, Superti-Furga et al. 1995). These mutations (two at amino acid position 380 and one at 375) were excluded as a cause of dwarfism in Dexter cattle because of the lack of difference between the sequence in a "bull-dog" calf and a non-carrier Dexter cow. This indicated that the mutation causing achondroplastic dwarfism in man is not the same as that causing dwarfism in Dexter cattle. It is interesting to note that both amino acid substitutions resulting in ACH occur at the conserved glycine residues. Changing the small, non-polar hydrophobic glycine to the positively charged hydrophilic arginine or to a large sulphur containing cysteine could have a significant effect on the structure of the TM region of FGFR3 protien. Understanding the molecular basis of the phenotype caused by the glycine-arginine and glycine-cysteine substitution in FGFR3 is most likely to provide insights into basic issues in bone growth and development in normal and diseased states and contribute to knowledge regarding structure and functional aspects of FGFRs in general.

6.4.3 Growth factors and the "bull-dog" phenotype.

The phenotypic peculiarities of homozygous mutant Dexter cattle, characterised by extreme shortening of limbs and craniofacial defects in the so called "bull-dog" appearance, suggests the involvement of the bones associated with cartilage templates and the intramembranous bones of the skull. FGFR3 cannot be completely excluded as the cause of Dexter dwarfism. A comparative study of the phenotype of the homozygous normal, heterozygous carrier and homozygous lethal in ACH would possibly indicate the role of FGFR3 in normal growth. In man, a heterozygous carrier (ACH/-) with one mutant gene of FGFR3 was half the size of a homozygous normal. The homozygous lethal (ACH/ACH) with two mutant genes of FGFR3 is half the size of the heterozygous carrier at infancy (Pauli et al. 1983). Another study in mice with homozygous FGFR3 deletions showed overgrown long bones (Muenke and Schell

1995), indicating a regulatory role of FGFR3 in growth and development. Table 6.6 summarizes the phenotypes in patients with ACH and different doses of mutant FGFR3. Similar mechanisms in the phenotypes may be applicable in Dexter cattle dwarfism. However, the cause of disproportionate bone growth especially in long bones and irregularities in facial bones in Dexter cattle has to be investigated.

The "bull-dog" phenotype has close similarity with that of Thanatophoric dysplasia in man (Tavormina et al. 1995), characterised by severe shortening of long bones in association with craniosynostosis, also suggesting an involvement of FGFR3. Another possible candidate gene is FGFR2 which is known to cause craniosynostosis and is also known to be expressed in the developing limb bud (Reardon et al. 1994). A Crouzon-like syndrome in mouse is known to be caused by insertional mutation in the FGF4 region (Carlton and Evans 1995). Thus a possible role of other FGFs and their receptors can be suggested in Dexter dwarfism.

The basic mechanism of the involvement of any of the growth factors or receptors in development seems similar. The mutant forms of growth factor receptors may alter the conformational change and may not migrate properly through the cell membrane, which might inhibit dimerization of the wild type form of the receptor with itself, with the other FGFRs or with both. Alternatively, disruption of the structure of the transmembrane domain may prevent the activation of the tyrosine kinase function, especially as the glycine to arginine mutation involves a highly conserved status (Ornitz and Leder 1992) and introduces a charged residue into a highly hydrophobic domain of the receptor.

6.4.4 Future experiments needed to exclude/include FGFR3.

Linkage analysis on informative pedigrees is the appropriate approach for locating a disease gene. However, many non-paternities were identified in the Dexter families available for this study. So the correct families have to be identified and DNA samples collected from these families. Also the diagnosis of carrier/non-carrier status of the defect has to be correctly confirmed. Identification of a polymorphism in the FGFR3 gene could be carried out either by DGGE or SSCP analysis. Once a polymorphism is detected, this can be used for linkage analysis in the informative

Table 6.6 - FGFR3 mutations and phenotype

	Heterozygote knock out mouse	Homozygote knock out mouse	Human Achondroplasia		Thanatophoric dysplasia	Human Hypochondroplasia		Normal
FGFR3	Knock out/Normal	Absent/Absent	Classical TM/Classical TM Mutant Mutant	Classical TM Mutant/Normal	TD/Normal	HCH/Normal	ACH/HCH	Normal/Normal
Aminoacid Changes	-/+	-/-	Gly380Arg/Gly380Arg Gly375Cys/Gly375Cys	Gly380Arg/Gly Gly375Cys/Gly	Arg248Cys/Arg Leu650Glu/Leu Ser371/Cys/Ser	Arg540Leu/Arg	Gly380Arg/Arg540Lys	Gly380/Gly380 Gly375/Gly375
Phenotype	Normal phenotype	Long bones	Very short bones	Short bones	Very short bones like ACH/ACH	Shorter bones, not as short as ACH/ACH	Severe dwarf individual	Normal phenotype
Hypothesised action of genotype	Normal signal present	No stop signal present	Very dominant premature stop	Mild dominant premature stop	Very dominant action causing stoppage of bone growth	Mild dominant premature stop	Dominant premature stop	Normal signal present
References	Muenke & Schell 1995	Muenke & Schell 1995	Shiang et al 1994 Rousseau et al 1995	Shiang et al 1994	Tavormina et al 1994	Bellus et al 1995	Bellus et al 1995	

pedigrees.

If FGFR3 is linked with the defect, the entire bovine DNA sequence for FGFR3 will be required. This could be done by screening the bovine cDNA library (in which FGFR3 was present) with the complete human cDNA probe. Sequencing of the equivalent DNA from the "bull-dog" calf should then identify any inherited change in the FGFR3 protein which may be the cause of Dexter dwarfism.

If no linkage of FGFR3 to the Dexter dwarf phenotype is found the possibility of FGFR2 as a candidate gene can be verified by linkage analysis, either by finding an informative polymorphism within FGFR2, or with markers selected from the bovine map.

If these also prove not to be the affected genes, the FGFs themselves are the next most likely candidate genes. Then a whole genome approach is required. This can be done by typing the Dexter families with a panel of markers available from the genome map and analysis of the segregation in families to identify the chromosomal region containing the gene responsible for the mutation. The corresponding region of the human genome can then be examined for obvious candidate genes for Dexter dwarfism. If none are identified a laborious positional cloning process involving YACs will have to be undertaken.

CHAPTER 7
GENERAL DISCUSSION

In the recent past, genetic improvement of domestic animal production has made tremendous strides with the advent of biotechnological tools. Animal breeders were quick to exploit the new biotechnological innovations for the propagation of economically beneficial and genetically stable germplasm. Artificial insemination and embryo transfer technology have ushered in dramatic improvements in milk production and carcass quality through a massive influx into the gene pool of genes of superior sires and a concomitant expansion in the number of superior offspring resulting from the crossing of elite females with progeny tested males.

Livestock improvement is a complex and time-consuming process and is based on sound, well-established principles and practices of animal breeding. The rapid technological expansion in molecular genetics is bound to revolutionize animal breeding and improvement programmes which are likely to emphasise procedures such as marker-assisted selection towards desirable production traits. The availability of the bovine genome map is a significant milestone as it provides a basis to discover genes controlling traits. The identification and molecular characterization of DNA markers for genes controlling particular traits will undoubtedly facilitate the selection of individuals carrying the appropriate alleles with the potential to improve several loci simultaneously. The mapping of the economically important trait loci in cattle will require a linkage map of the entire bovine genome with about 150-200 evenly spaced markers that are polymorphic in the appropriate populations. The information on the location of markers in the bovine genome map will have a direct application on improving the genetic merit of cattle breeds in developing countries.

In the industrialized regions of the world, genome mapping has received considerable attention in agricultural and animal breeding research. The main thrust will be in investigations aimed at revealing relationships between genes and production traits and disease susceptibility and resistance characteristics. Several

gene related deficiencies and disorders have been identified in man and animals.

Genetic markers are also useful for introgressing a particular gene, related to a trait of interest between populations to increase their overall performance. Thus marker assisted introgression is especially important in improving the chosen production traits and at the same time retaining their innate resistance characteristics and adaptability. The study by FAO to introgress the gene causing trypano-tolerence from N'Dama cattle to trypano-sensitive tropical breeds is an example in this context (Beckmann and Soller 1987).

Apart from marker assisted selection and introgression, DNA based methodologies have defined advantages in conventional breeding practices, eg: in parentage verification, line and hybrid purity testing, exploration of homologous loci in other species or genera and germplasm evaluation. In progeny testing and MAS, paternity of each individual included in the test is important for effective and accurate determination of the "proof" of the sire. Identity of the bulls included in the progeny test are currently confirmed by the conventional blood typing method. The use of DNA markers will be introduced in the near future.

There are several disadvantages in the use of blood typing in routine parentage testing. The absence of monovalent antisera specific for certain breeds precludes the use of reagents prepared from one breed to another. Even though comparison tests are carried out between laboratories for standardization of typing reagents, new reagents if detected in zebu breeds, often cause confusion. The battery of reagents prepared in the same laboratory at different times may differ in reactivity so that the samples genotyped twice might show addition/missing of certain specificities.

However, with the introduction of DNA markers, RFLP, mini, midi and microsatellite have been applied in different laboratories. The polymerase chain reaction improves the analysis of polymorphism based on the length variation in tandemly repeated nucleotide sequences, the so-called VNTRs.

In this thesis, the merits and shortcomings of RFLP and minisatellites for genome mapping have been emphasized. Although PCR has improved the efficiency of detection of RFLPs, the occurrence of few alleles with low

heterozygosity and PIC and the need to use large numbers have restricted their usefulness in a routine procedure. Southern blotting is laborious, time-consuming and cost-prohibitive. The minisatellite approach is equally laborious and expensive. The preferential telomeric location of minisatellite sequences in cattle has rendered them unsuitable for linkage analysis. Also the minisatellite repeats are too large (20-25kb) to be amplified efficiently by PCR.

In distinct contrast to the RFLP and minisatellite loci, the microsatellites provide a better and more efficient system of genome analysis because of their high PIC and heterozygosity value and amenability to PCR amplification and automation. The presence of narrow size ranges of alleles, the relative reproducibility and ease of detection of alleles and the possibility of multiplexing improve the merits of microsatellites. Small amounts of DNA are sufficient for the tests. Therefore, microsatellite markers are now being tested in several blood group testing laboratories as an alternate approach for parentage verification. In the UK, bovine microsatellite markers have been successfully used as DNA evidence in courts. Their potential use in evolutionary studies of cattle breeds has been described (MacHugh et al. 1994, Loftus et al. 1994).

In this thesis, the usefulness of microsatellite markers was investigated in three separate studies, namely, parentage verification, determination of phylogenetic relationship between selected cattle breeds and mapping a lethal defect in Dexter cattle.

Five markers, CYP21, DRB3, FSHB, ETH131 and HEL6, were used in a parentage verification study of 275 cattle belonging to 15 European breeds. Two markers were closely-linked and the others located on different chromosomes. All five markers were highly polymorphic and this accounted for their high efficiency in parentage verification. It was noteworthy that a significantly greater number of alleles was observed for each marker than were reported in several earlier studies with the same markers (Fries et al. 1990, Kemp and Teale 1991, Ellegren et al. 1993, Steffan et al. 1993, Kaukinen and Varvio 1993). This is perhaps attributable to the fact that the breeds used in the present study consisted of a large number of unrelated individuals thus providing a broader genetic base.

A new approach for the statistical evaluation of probability of random sire exclusion (PRASE) was made taking into consideration the deviation from HWE and linkage between markers and using observed genotype frequencies. Calculated together, the two linked markers DRB3 and CYP21 gave an average PRASE of 0.88 in all breeds, the range being 0.75 - 0.96. With the inclusion of each additional marker, the PRASE was increased to 0.99 for all the breeds. This was a similar value to that observed with the "StockMarks" kit of Perkin Elmer containing 11 microsatellite markers. Using this kit the observed probability of exclusion of a random incorrect sire was low for each marker, but when all the 11 markers were used the value approximated to 0.99. However, the ABI data is based only on the Holstein Friesian breed of cattle. The comparison of PRASE in the present study with that reported in the studies using "StockMarks" kit clearly indicates that the polymorphic nature of markers is inversely proportional to the numbers used in parentage verification to obtain the maximum probability of exclusion.

Multiplexing or simultaneous analysis of all the markers in a gel lane is a recent innovation. In the present study, ABI 373 sequencer with five fluorescently labelled markers made it possible to analyse all the markers in a single gel lane. The narrow size range of these markers together with the colour effect of the different fluorescent tags helped in discrimination of the PCR products. To obtain efficient co-amplification the reaction condition for each marker was carefully adjusted. Also it was observed that the PCR products analysed on a single gel lane could be considerably improved by multiple loading.

Several investigators, notably Buchanan et al. (1994) and MacHugh et al. (1994), suggested that the high polymorphic nature of microsatellites may render them more suitable as indicators of evolutionary relationships within species than other classes of markers. Also features such as their conservation between breeds were thought to be good measures for use in phylogenetic studies. In the present study using five microsatellite markers a significant variation in allele frequency was observed between the 15 breeds of cattle. For instance, for the Dexter, Belted Galloway and Boran breeds, only a few alleles were detected for all the five

markers as compared to the Friesian and Limousin breeds. Valdes et al. (1993) reported a direct relationship between allelic length difference and evolutionary divergence.

In the present study, phylogenetic trees were constructed using three well established methods of calculation of genetic distance: Nei's, Cavalli's and Reynold's. The neighbor joining method was used to construct the tree topology for each marker. However, none of these approaches gave conclusive results. The exception was a consistent grouping of breeds Jersey & Guernsey, South Devon & Belted Galloway and Friesian, Limousin & Blonde Aquitaine in all the three topologies.

Crandall and Templeton (1993) offered the opinion that construction of intraspecies phylogenies is difficult for cattle populations. They assumed that a better phylogenetic picture may emerge if the markers used represented the entire genome. A relevant comment in this context (Nei and Roy Choudhary 1972) is that for estimation of genetic distance between breeds the use of a greater number of loci rather than a good number of individuals per locus might be rewarding.

At the time of the present study in which a relationship or otherwise between the genes causing the bovine disproportionate dwarfism ("bull-dog" calf) and human achondroplasia was sought, published mapping data had shown that at least three point mutations in the TM region of FGFR3 were the cause of the human dwarfism and that the sequence associated with it was located on chromosome 4, which is syntenic to bovine chromosome 6. Because of the insufficient number of polymorphic markers on bovine chromosome 6, markers on the TM region of human FGFR3 were employed for linkage analysis. However, in preliminary studies no conclusive results emerged.

When microsatellite markers on bovine chromosome 6 became available some Dexter families were typed. Although these markers were informative in the chosen families, they failed to establish a linkage. However, pedigree errors in some families were detected. Results of linkage analysis showed a negative lod score for closely linked markers, a finding that indicated that the families studied may not be useful for further linkage analysis.

Subsequent studies using human markers were marred by the problem of contamination with human DNA. As this problem seemed to occur only with the use of human FGFR3 specific primers, the determination of the sequence in the TM region of bovine FGFR3 molecule by a different method was necessary. The application of touch down PCR in which longer PCR primers from human FGFR3 sequence were used resulted in amplification of the desired bovine specific product. Sequencing of this 182bp PCR product revealed a highly GC-rich sequence comparable to the sequence of human FGFR3. However, there were marked differences with the published sequences in human, mouse and chicken FGFR3 molecules. This finding clearly suggested that the gene associated with bovine disproportionate dwarfism is not the equivalent of the human ACH gene.

The data presented in this work amply support the thesis that microsatellites are valuable tools for parentage verification and individual identification studies. They also have a potential application in establishing phylogenetic relationships within species. Their use will enable identification of sequences in the genome associated with production traits. Defects in such sequences may be the cause of inherited syndromes as in the Dexter "bull-dog" condition. Further sequential analysis of the bovine genome and the identification of polymorphic markers will be useful in such investigations. The laboratory techniques, typing methods and to a certain extent the markers used are the same for parentage testing, genome mapping, QTL selection and phylogeny studies. Therefore, a laboratory set up for one of these procedures can be easily used for the others.

Biotechnological tools such as microsatellite markers of proven value have their potential use in several developing countries with a good infrastructure and established background animal breeding data. India has the largest cattle and buffalo populations in the world (15 and 55 percent respectively) and is also the second largest milk producer. However, it has remained economically poor. The animal protein intake *per capita* is far below the world average. Over the decades, there has been a gradual deterioration in the productivity of indigenous milk breeds, in particular, because of an uncontrolled increase in the number of animals, indiscriminate breeding practices and the lack of organised efforts aimed at a

systematic and sustained improvement of genetically superior indigenous breeds.

Elsewhere, two major tools, intensive selection and rigid culling, operating in concert have helped the development of superior breeds. On the other hand, India has been witnessing the paradox of scientists having the freedom to select stocks of proven economically beneficial parameters and achieving significant gains, but the country as a whole lagging behind due the ban on culling substandard stock due to social constraints. The result is a significant increase in cattle and buffalo populations, but with no concomitant increase in population - related milk production and the generation of superior stock.

In India, cross breeding of selected indigenous milk breeds with a wide range of exotic dairy breeds of proven productive traits has been systematically applied. This programme of upgrading indigenous stock has been supported by systematic performance testing of heifers and progeny testing of the bulls. Artificial insemination and embryo transfer technologies have been widely applied to obtain generations of genetically superior crossbreds. However, the role of proven sires for sustaining genetic gains is still being assessed and needs greater attention. With the advent of multiple ovulation and embryo transfer technology (MOET), performance testing of female stock and full exploitation of the genetic potential of their dams are increasingly evident. However, progeny and performance test are time-consuming. Identification of QTL and a programme of marker-assisted selection could be implemented to achieve improvement by faster, efficient and cost effective methods. Moreover, with the development of superior crossbreds and their wide distribution in the country indiscriminate breeding and illegal practices will have to be identified and eliminated. Parentage verification and phylogenetic studies may have to be conducted for effective sire evaluation and preservation of genetic gains and germplasm.

One basic limitation of applying biotechnology to animal genetic improvement is the limited knowledge of the number and chromosomal location of genes that influence quantitative traits. Recently the bovine genome map is sufficiently covered with polymorphic markers that it can be used for identifying a QTL. In future, it will be possible to identify markers that are linked to QTL

and this information will be important for use in MAS programmes in developing countries where selection can be implemented on the basis of markers known to be linked to QTL.

The potential applications of marker based technology in both industrialized and developing regions of the world are manifold. Any programme of using a specific and selective tool such as microsatellite markers should take into consideration basic lacunae such as the limited knowledge of the molecular basis of their occurrence, and the small numbers and chromosomal location of known genes that influence quantitative traits. It does not need to be emphasised that a greater number of markers, particularly those that are linked to QTL, need to be identified before this development will greatly assist implementation of marker assisted selection and introgression programmes. This will be particularly beneficial to countries such as India where selection can be implemented only on the basis of such proven markers.

REFERENCES

Ali S., Muller C.R. and Epplen J.T. (1986) DNA fingerprinting by oligonucleotide probes specific for simple repeats. *Hum. Genet.* **74**, 239-243.

Allen R.W., Wallhermefehtel M. and Miller W.V. (1990) The application of restriction fragment length polymorphism mapping to parentage testing. *Transfusion.* **30**, 552-564.

Amigues Y., Brathwaite B., Pepin L. and Boscher M.Y. (1994) Use of microsatellites as an alternative to bloof typing in goat parentage control. *Anim. Genet.* **25**, 71.

Ammerman A.J. and Cavalli-Sforza L.L. (1984) *The neolithic transition and the genetics of populations in Europe.* Princeton University Press, Princeton, Nj.

Anderson S., DE Bruijn M.H.L., Coulson A.R., Eperon I.C., Sanger F. and Young I.G. (1982) Complete sequence of bovine mitochondrial DNA. *J. mol. Biol.* **156**, 683-717.

Anderson L., Bohme J., Rask L. and Peterson P.A. (1985) Genomic hybridization of bovine class II major histocompatibility genes: 1. Extensive polymorphism of DQα and DQβ genes. *Anim. Genet.* **17**, 95-98.

Andersen W.R. and Fairbanks D.J. (1990) Molecular markers: important tools for plant genetic resource characterization. *Diversity.* **6**, 51-53.

Armour A.J.L. and Jeffreys A.J. (1992) Biology and applications of human minisatellite loci. *Curr. Opin. Genet. Devpt.* **2**, 850-856.

Avise I.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A. and Saunders N.C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Sys.* **18**, 489-522.

Baker C.M.A. and Manwell C. (1991) Population genetics, molecular markers and gene conservation of bovine breeds. *In Cattle Genetic Resources.* (ed. C.G.Hickman), 221-304.

Balakrishnan V. and Sanghvi L.D. (1968) Distance between populations on the basis of attribute data. *Biometrics.* **24**, 859-865.

Barendse W., Armitage S.M., Kossarek L.M. et al. (1994) A genetic linkage map of the bovine genome. *Nature Genet.* **6**, 227-235.

Barker D., Schafer M. and White R. (1984) Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell.* **36**, 131-138.

Baron B., Poirier C., Simon-Chazottes D., Barnier C.and Guenet J.L. (1992) A new strategy useful for rapid identification of microsatellites from DNA libraries with large size inserts. *Nucl. Acids Res.* **20**, 3665-3669.

Basilico C. and Moscatelli D. (1992) The FGF family of growth factors and oncogenes. *Adv. Cancer Res.* **59**, 115.

Basrur P.K. (1980) Genetics in veterinary medicine. In *Scientific foundation of Veterinary medicine.* (ED. Phillipson A.T, Hall L.W, Prichard W.R) William Heinemann Medical Books Ltd. 393-413.

Bates G.P., Wainwright B.J., Williamson R. and Brown S.D.M. (1986) Microdissection of and microcloning from the short arm of human chromosome-2. *Mol. Cell. Biol.* **6**, 3826-3830.

Beckmann J.S and Soller M. (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.* **67**, 35-43.

Beckmann J.S. and Soller M. (1987) Molecular markers in the genetic improvement of farm animals. *Biotechnology.* **5**, 573-576.

Beckmann J.S. and Soller M. (1990) Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Biotechnology.* **8**, 930-932.

Beckmann J.S., Tomfohrde J., Barnes R.I. et al. (1993) A linkage map of human chromosome 15 with an average resolution of 2cM and containing 55 polymorphic microsatellites. *Hum. mol. Genet.* **2**, 2019-2030.

Beever J.E., George P.D., Fernado R.L., Stormont C.J. and Lewin H.A. (1990) Associations between genetic markers and growth and carcass traits in a paternal half-sib family. *J. Anim. Sci.* **68**, 337-344.

Bell G.I., Selby M.J.and Rutter W.J. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature.* **295**, 31-35.

Bellus G.A., McIntosh I., Anne Smith E. et al. (1995) A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nature Genet.* **10**, 357-359.

Bender W., Spierer P. and Hogness D.S. (1983) Chromosomal walking and

jumping to isolate DNA from the ACE and Rosy loci and the bithorax complex in drosophila melenogaster. *J. Mol. Biol.* **168**, 17-33.

Bhat P.P., Mishra B.P. and Bhat P.N. (1990) Polymorphism of mitochondrial DNA (mtDNA) in cattle and buffaloes. *Biochem. Genet.* **28,** 311-318.

Biggin M.D., Gibson T.J. and Hong G.F. (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Nat. Acad. Sci. USA*, **80**, 3963-3965.

Bishop M.D., Kappes S.M., Keele J.W., Stone R.t., Sunden S.L.F., Gregory H.A., Toldo S.S., Fries R., Grosz M.D., Yoo J. and Beattic C.W. (1994) A genetic linkage map for cattle. *Genetics*. **136**, 619-639.

Bodmer W.F., Bailey C.J., Bodmer J. et al (1987) Localisation of the gene for familial adenomatous polyposis on chromosome 5. *Nature*. **328**, 614-616.

Botstein D., White R.L., Skolnick M. and Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet*. **32,** 314-331.

Bowcock A.M., Ruis-Linares A., Tomfohrde J., Minch E., Kidd J.R. and Cavalli-Sforza. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*. **368**, 455-457.

Bradley D.G., MacHugh D.E., Loftus R.T., Sow R.S. and Hoste C.H. (1994) Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Anim. Genet*. **25**, 7-12.

Brown W.M., George M. and Wilson A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*. **76**, 1967-1971.

Buchanan F.C., Adams L.J., Littlejohn R.P., Maddox J.F. and Crawford A.M. (1994) Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics*. **22**, 397-403.

Buitkamp J., Zischler H., Epplen J.T. and Geldermann H. (1991) DNA fingerprinting in cattle using oligonucleotide probes. *Anim. Genet*. **22**, 137-146.

Bulfield G., Siller W.G., Wight-Pal. and Moore K.J (1984) X-chromosome-linked muscular dystrophy (mdx) in the mouse. *Proc. Natl. Acad. Sci. USA*. **81**, 1189-1192.

Burgess W.H. and Maciag T. (1989) The heparin-binding (fibroblast) growth factor family of proteins. *Ann. Rev. Biochem.* **58**, 575-606.

Burke T. and Bruford M.W. (1987) DNA fingerprinting in birds. *Nature.* **327**, 149-152.

Burrus L.W., Zuber M.E., Lueddecke B.A. and Olwin B.B. (1992) Identification of a cysteine-rich receptor for fibroblast growth-factors. *Mol. Cell. Biol.* **12**, 5600-5609.

Callen D.F., Thompson A.D., Shen Y., Phillips H.A., Richards R.I. and Southerland G.R. (1993) Incidence and origin of null alleles in the (AC) N microsatellite markers. *Am. J. Hum. Genet.* **52**, 922-927.

Carlton M. and Evans M. (1995) A Crouzon like craniofacial dysmorphology in the mouse is caused by insertional mutation in the FGF4 region. *Genet. Res.* **65**, 234.

Carrano A.V., Lamerdin J., Ashworth L.K. et al. (1989) A high-resolution fluorescence based semi-automated method for DNA fingerprinting. *Genomics.* **4**, 129-136.

Caskey C.T., Edwards A. and Hammond H.A. (1989) DNA: the history and future use in forensic analysis. *Proceedings of the 1989 International Symposium of the Forensic application of DNA analysis. Quantico. Virginia.* 3-9.

Cavalli-Sforza L.L and Edwards A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *Evolution.* **32**, 550-570. (also Amer J. Hum. Genet. 19, 233-257.

Ceppellini R., Siniscalco M. and Smith C.A.B. (1956) The estimation of gene frequencies in a random-mating population. *Ann. Hum. Genet.* **20**, 97-115.

Chakraborty R., Meagher T.R and Smouse P.E. (1988) Parentage analysis with genetic markers in natural populations. 1. The expected proportion of offspring with unambigous paternity. *Genetics.* **118**, 527-536.

Chamberlain J.S., Gibbs R.A., Ranier J.E. and Caskey C.T. (1990) Multiplex PCR for the diagnosis of Duchenne Muscular Dystrophy. *PCR Protocols: A guide to methods and applications.* 272-281.

Chung B., Matheson K.J. and Miller W.L. (1986) Structure of a bovine gene for P-450C21 (steroid 21 hydroxylase) defines a novel cytochrome P-450 gene family. *Proc. Natl. Acad. Sci. USA* **83,** 4243-4247.

Cooper D.N., Smith B.A., Booke H.J., Niemann S. and Schmitke J. (1985) An estimate of unique sequence heterozygosity in the human genome. *Hum. Genet.* **69**, 201-205.

Cornall R.J., Aitman T.J., Hearne C.M. and Todd J.A. (1991) The generation of

a library of PCR-analysed microsatellite variants for genetic mapping of the mouse genome. *Genomics.* **10**, 874-881.

Cotton R.G.H., Rodrigues N.R. and Campbell R.D. (1988) Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA.* **85**, 4397-4401.

Crandall K.A. and Templeton A.R. (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics.* **134**, 959-969.

Crew F.A.E. (1924) The bull-dog calf: a contribution to the study of achondroplasia. *Proc. Royal Society of Medicine (Comparative medicine).* **17**, 39-58.

Cunningham E.P. (1990) Quantitative approaches to animal improvement. *Proceedings of the 4th World Congress on Genetics applied to livestock production. Edinburgh.* **XV1**, 4-15.

Cunningham E.P. (1992) In *Selected Issues in Livestock Industry Development* (The World Band, Washinghton, DC) 1-12.

Cunningham P., Loftus R.T., MacHugh D.E. and Bradley D.G. (1994) Molecular evolution of African, European and Asian cattle. *Proceedings of the 5th World Congress on Genetics applied to livestock production. Guelph, Ontario, Canada.* **XX1**, 86-89.

Dain A.R., Tucker E.M., Donker R.A. and Clarke S.W. (1984) Chromosome mapping in cattle using mouse myeloma/calf lymph node cell hybridomas. *Biochem. Genet.* **22**, 429-439.

Dallas J.F. (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome.* **3**, 452-456.

Dausset J.H., Cann H., Cohen D., Lathrop M., Lalouel M. and White R. (1990) Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative mapping of the human genome. *Genomics.* **6**, 575-577.

Davies K.E., Jackson J., Williamson R., Harper P.S., Ball S., Sarfarazi M., Meredith L. and Fey G. (1983) Linkage analysis of myotonic dystrophy and sequences on chromosome 19 using a cloned complement 3 gene probe. *J. Med. Genet.* **20**, 259-263.

Davis W., Harbitz I., Fries R., Stranziger G. and Hauge J.G. (1988) Porcine malignant hyperthermia carrier detection and chromosomal assignment using a linked

probe. *Anim. Genet.* **19**, 203-212.

Devlin B., Resch N. and Roeder K. (1990) No excess of homozygosity at loci used for DNA fingerprinting. *Science.* **249**, 1416-1420.

Diehl S.R., Ziegle J., Buck G.A., Revnolds T.R. and Weber J.L. (1990) Automated genotyping of human DNA polymorphisms. *Amer. J. Hum. Genet.* **47**, A177.

Dionne C.A., Kaplan R., Seuanez H., O'Brien S.J. and Jaye M. (1990) Chromosome assignment by the polymerase chain reaction techniques: assignment of the oncogene FGF5 to human chromosome 4. *BioTechniques.* **8**, 190-194.

Donis-Keller H., Green P., Helms C. et al (1987) A genetic linkage map of the human genome. *Cell.* **51**, 319-337.

Don R.H., Cox P.T., Wainwright B.J., Baker K. and Mattick J.S. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucl. Acids Res.* **19**, 4008.

Dunn L.C. and Bennet D. (1967) Sex differences in recombination of linked genes in animals. *Genet. Res.* **9**, 211-220.

Eck R.V. and Dayhoff M.O. (1966) Atlas of protein sequences and structure. *National Biomedical Research Organisation.*

Economou E.P., Bergen A., Warren A.C. and Antonarakis S.E. (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc. Natl. Acad. Sci. USA.* **87**, 2951-2954.

Edwards A.W.F. and Cavalli-Sforza L.L. (1963) The reconstruction of evolution. *Ann. Hum. Genet.* **27**, 105.

Edwards A.W.F. and Cavalli-Sforza L.L. (1964) Reconstruction of evolutionary trees. *Phenetic and Phylogenetic classification.* ed. Heywood V.H. and McNeill J. Syst. Assn. Publ. London. 67-76.

Edwards A., Civitello A., Hammond H.A. and Caskey C.T. (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Amer. J. Hum. Genet.* **49**, 746-756.

Edwards A., Hammond H.A., Jin L., Caskey C.T. and Chakraborty R. (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics.* **12**, 241-253.

Ellegren H., Davis C.J. and Andersson L. (1993) Strong association between

polymorphisms in and intronic microsatellite and in the coding sequence of the BoLA-DRB3 gene implications for microsatellite stability and PCR-based DRB3 typing. *Anim. Genet.* **24**, 269-275.

Erlebacher A., Filvaroff E.H., Giteiman S.E and Derynck R. (1995) Toward a molecular understanding of skeletal development. *Cell.* **80**, 371-378.

Evett I.W. and Gill P. (1991) A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigation. *Electrophoresis.* **12**, 226-230.

Fallon J.F., Lopez A., Ros M.A., Savage M.P., Olwin B.B. and Simandl B.K. (1994) Apical ectodermal ridge growth signal for chick limb development. *Science.* **264,** 104-107.

FAO (1993) In *World Watch List for Domestic Animal Diversity* (ed. by R.T Loftus and B. Scherf), 1st edn. FAO, Rome.

Felius M. (1985) *Genus Bos: Cattle breeds of the world*, Merck & Co INC, USA. pp. 193-226.

Felsenstein J. (1981) Evolutionary trees from DNA sequences - a maximum likelihood approach. *J. Mol. Evol.* **17**, 368-376.

Felsenstein J. (1985) Phylogenies from gene frequencies: a statistical problem. *Syst. Zool.* **34**, 300-311.

Felsenstein J. (1993) *Phylogeny inference package, version 3.5c.* Department of Genetics. Univ. Washington, Seattle.

Fincham J.R.S. (1983) *Genetics.* John Wright and Sons Ltd. Stonebridge Press Bristol. 823-825.

Folkman J. and Klagsbrun M. (1987) Angiogenic factors. *Science.* **235**, 442-447.

Fowler S.J., Gill P., Werret D.J.and Higgs D.R. (1988) Individual specific DNA fingerprints from a hypervariable region probe: Alpha Globin 3'HVR. *Hum. Genet.* **79**, 142-146.

Francomano C.A., Ortiz de Luna R.I., Hefferon T.W. et al. (1994) Localisation of the achondroplasia gene to the distal 2.5 Mb of human chromosome 4p. *Hum. Mol. Genet.*
3, 787-792.

Fraser D.C., Craigmile S. and Russell G.C. (1994) Molecular cloning and sequencing of a cattle DRA cDNA. *Immunogenetics.* **40**, 311.

Fries R., Hediger R and Stratzinger G. (1986) Tenative chromosomal lacalization of the bovine histocompatibility complex by in situ hybridisation. *Anim. Genet.* **17,** 287-294.

Fries R., Hediger R and Stratzinger G. (1988) The loci for parathyroid hormone and beta lactoglobulin are closely linked and map to chromosome 15 in cattle. *Genomics.* **3**, 302-307.

Fries R., Beckmann J.S., Georges M., Soller M. and Womack J. (1989) The bovine gene map. *Anim. Genet.* **20**, 3-29.

Fries R., Eggan A and Stratzinger G. (1990) The bovine genome contains polymorphic microsatellites. *Genomics.* **8**, 403-406.

Fries R., Threadgill D.W., Hediger R., Gunavwardana A., Blessing M., Jorcano J.L., Stranzinger G. and Womack J.E. (1991) Mapping of the bovine cytokeratin sequence to four different sites on three chromosomes. *Cytogenet. Cell Genet.* **57**, 135-141.

Fries R. (1993) Mapping the bovine genome: methodological aspects and strategy. *Anim. Genet.* **24**, 111-116.

Fries R., Eggen A., James E. and Womack. (1993) The bovine genome map. *Mammalian Genome.* **4**, 405-428.

Fujii J., Otsu K., Zorzato F., DeLeon S., Khanna V.K., Weiler J.E., O'Brien P.J. and Maclennan D.H. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science.* **253**, 448-451.

Gahne B. (1961) Studies of transferrins in serum and milk of Swedish cattle. *Anim. Prod.* **3**, 135-145.

Gallagher D.S., Threadgill D.W., Jackson A.P., Perham P. and Womack J.E. (1991) Somatic cell mapping of bovine clathrin light chain genes: identification of a new syntenic group. *Cytogenet. Cell Genet.* **56**, 154-156.

Gallagher D.S. and Womack J.E. (1992) Chromosome conservation in the bovidae. *J. Hered.* **83**, 287-298.

Gardner E.J., Simmons M.J. and Snustad D.P. (1991) *Principles of genetics.* 8th edn. New York, Chichester, Wiley.

Geldermann H. (1975) Investigations on inheritance of quantitative characters in animals by gene markers. 1. methods. *Theor. Appl. Genet.* **46**, 319-330.

Geldermann H. (1976) Investigations on inheritance of quantitative characters in

animals by gene markers. II. expected effects. *Theor. Appl. Genet.* **47**, 1-4.

Geldermann H., Pieper U. and Roth B. (1985) Effects of marked chromosome sections on milk performance in cattle. *Theor. Appl. Genet.* **70**, 138-146.

Georges M., Swillens S., Bouquet Y., Lequarre A.S. and Hanset R. (1987) Genetic linkage between the bovine plasma protease inhibitor 2 (Pi-2) and S blood group loci. *Anim. Genet.* **18**, 311-316.

Georges M., Hilbert P., Lequarre A., Lec lerc V., Hanset R. and Vassart G. (1988a) Use of DNA bar codes to resolve a canine paternity dispute. *J. Amer. Vet. Med. Assoc.* 193-199.

Georges M., Lequarre A., Castelli M., Hanset R. and Vassart G. (1988b) DNA fingerprinting in domestic animals using four different minisatellite probes. *Cytogenet. Cell Genet.* **47**, 127-131.

Georges M., Lathrop M., Hilbert P., Marcottr A., Schwers A., Swillens S., Vassart G. and Hanset R. (1990a) On the use of DNA fingerprinting for linkage studies in cattle. *Genomics.* **6**, 461-474.

Georges M., Lathrop M., Bouquet Y et al. (1990b) Linkage relationships among 20 genetic markers in cattle. Evidence for linkage between two pairs of blood group systems: B-Z and S-F/V respectively. *Anim. Genet.* **21**, 95-105.

Georges M. (1991) Hypervariable minisatellites and their use in animal breeding. *Gene-Mapping Techniques and Applications.* (ed. by Schook L.B, Lewin H.A, McLaren D.G.) 89-112.

Georges M., Dietz A.B., Mishra A., Nielsen D., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Leipold H., Womack J.E. and Lathrop M. (1993a) Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proc. Natl. Acad. Sci. USA.* **90**, 1058-1062.

Georges M., Drinkwater R., King T., Mishra A., Moore S.S., Neilsen D., Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E. and Hetzel J. (1993b) Microsatellite mapping of a gene affecting horn development in Bos taurus. *Nature Genet.* **4**, 206-210.

Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A.T., Sargent L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E. and Hoeschele I. (1994) Using a complete microsatellite map and the grand-daughter design to locate polygenes controlling milk production. *Proc. of the 5th World Congress on Genetics Applied to Livestock Production. Guelph, Ontario, Canada.* 81-85.

Georges M., Nielsen D., Mackinnon M., Mishra A., Okimoto R., Pasquino A.T.,

Sargeant L.S., Sorensen A., Steele M.R., Zhao X., Womack J.E. and Hoeschele I. (1995) Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics.* **139**, 907-920.

Gill P., Jeffreys A.J. and Werrett D.J. (1985) Forensic application of DNA fingerprints. *Nature.* **318**, 577.

Gill P. and Werrett D.J. (1987) Exclusion of a man charged with murder by DNA fingerprinting. *Forensic Sci. Intern.* **35**, 145-148.

Gill P., Sullivan K. and Werrett D.J. (1990) The analysis of hypervariable DNA profiles: problems associated with the objective determination of the probability of a match. *Hum. Genet.* **85**, 75-79.

Givol D. and Yayon A. (1992) Complexity of FGF receptors - genetic-basis for structural diversity and functional spwcificity. *FASEB J.* **6**, 3362-3369.

Glaser R.L., Thomas G.H., Siegfried E., Elgin S.C.R. and Lis J.T. (1990) Optimal heat induced expression of the Drosophila hsp26 gene requires a promoted sequence containing (CT)n.(GA)n repeats. *J. Mol. Biol.* **211**, 751-761.

Glowatzki-Mullis M.L. and Fries R. (1994) Parentage control in cattle by genotyping microsatellites. *Adv. Forensic Haemogenet.* **5**, 207-209.

Glowatzki-Mullis M. L., Gaillard C., Wigger G. and Fries R. (1995) Microsatellite-based parentage control in cattle. *Anim. Genet.* **26**, 7-12.

Gonyon D.S., Mather R.E., Hines H.C., Haenlein G.F.W., Arave C.W. and Gaunt S.N. (1987) Associations of bovine blood and milk polymorphisms with lactation traits: holsteins. *J. Dairy Sci.* **70**, 2585-2598

Goodbourn S.E.Y., Higgs D.R., Clegg J.B. and Weatherall D.J. (1983) Molecular basis of length polymorphism in the human zeta-globin gene complex. *Proc. Natl. Acad. Sci. USA.* **80**, 5022-5026.

Gray J.W., Dean P.N., Fuscoe J.C., Peters D.C., Trask B.J., Vandenengh G.J. and Vandilla M.A. (1987) High speed chromosome sorting. *Science.* **238**, 323-329.

Greene H.J., Leipold H., Huston K., Noordsy J.L. and Dennis S.M. (1973) Congenital defects in cattle. *Irish Vet. J.* **27**, 37.

Grodzicker T., Williams J., Sharp P. and Sambrook J. (1974) Physical mapping of temperature sensitive mutations of adenoviruses. In *Symposia on Quantitative Biology: Physical mapping of temperature sensitive mutations of adenoviruses. Cold spring harbor* **39**, 439-446.

Grosclaude F., Francois D. and Wimitzky M. (1990) Evidence for absence of linkage between the B and Z as well as between the F and S systems of cattle blood groups. *Anim. Genet.* **21**, 427-429.

Gross D.S. and Garrard W.T. (1986) The ubiquitous potential Z forming sequence of eukaryotes, (dT-dG)n.(dC-dA)n is not detectable in the genomes of eubacteria, archaebacteria or mitochondria. *Mol. Cell. Biol.* **6**, 3010-3013.

Guo S.W and Thompson E.A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. **48**, 361-372.

Gusella J.F., Keys C., Varsanyi-Breiner A., Kao F.T., Joens C., Puck T.T. and Housman D. (1980) Isolation and localisation of DNA segments from specific human chromosome. *Proc. Natl. Acad. Sci. USA.* **77**, 2829-2833.

Gusella J.F., Wexler N.S., Conneally P.M., Naylor S.L., Anderson M.A., Tanzi R.E., Watkins P.C., Ottina K., Wallace M.R., Sakaguchi A.Y., Young A.B., Shoulson I., Bonilla E. and Martin J.B. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature.* **306**, 234-238.

Gustavsson I. (1969) Cytogenetics distribution and phenotypic effects of a translocation in Swedish cattle. *Hereditas.* **63**, 68-169.

Haberfeld A., Cahaner A., Yoffe O., Plotsky Y. and Hillel J. (1991) DNA fingerprints of farm animals generated by microsatellite and minisatellite DNA probes. *Anim. Genet.* **22**, 299-305.

Haberfeld A., Kalay D., Weisberger P., Gal O. and Hillel J. (1993) Application of multilocus molecular markers in cattle breeding. 1.Minisatellites and Microsatellites. *J. Dairy. Sci.* **76**, 645-652.

Haenlein G.F.W., Gonyon D.S., Mather R.E. and Hines H.C. (1987) Associations of bovine blood and milk polymorphisms with lactation traits: Guernseys. *J. Dairy Sci.* **70**, 2599-2609.

Hall J.G. (1959) Blood grouping in domestic animals. *Vet. Rec.* **71**, 1062-1070.

Hall B.D. and Spranger J. (1979) Hypochondroplasia: clinical and radiological aspects in 39 cases. *Radiology.* **133**, 96-100.

Hall J.G. and Ross D.S. (1981) Evidence for the presence of an additional allele in the F system of British Friesian cattle blood. *Anim. Blood Grps. Biochem. Genet.* **12**, 229-240.

Hallerman E.M., Kashi Y., Nave A., Beckmann J.S. and Soller M. (1986) Restriction fragment length polymorphisms in dairy cattle and their utilization in

genetic improvement. *Review of Animal Production.* **22**, 31-37.

Hamada H., Petrino M.G. and Kakunaga T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci.USA.* **79**, 6465-6469.

Hamada H., Seidman M., Howard B.H. and Gorman C.M. (1984a) Enhanced gene expression by the poly (dT-dG) poly (dC-dA) sequence. *Mol. Cell Biol.* **4**, 2622-2630.

Hamada H., Petrino M.G., Kakunaga T., Seidman M. and Stollar B.D. (1984b) Characterisation of genomic poly (dT-dG) poly (dC-dA) sequence: structure, organisation and conformation. *Mol. Cell Biol.* **4**, 2610-2621.

Hauge X.Y. and Litt M. (1993) A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum. Mol. Genet.* **2**, 411-415.

Hawker K.J. and Billadello J.J. (1993) A multiplex polymerase chain reaction colony miniprep. *BioTechniques.* **14**, 764-765.

Hayasaka K., Kawamoto Y., Shotake T. and Nozawa K. (1986) Probability fo paternity exclusion and the number of loci needed to determine the fathers in a troop of macaques. *Primates.* **27**, 103-114.

Hazan J., Dubay C., Pankowiak M.P., Becuwe N. and Weissenbach J. (1992) A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. *Genomics.* **12**, 183-189.

Hecht J.T., Herrera C.A., Greenhaw G.A., Francomano C.A., Bellus G.A. and Blanton S.H. (1995) Confirmatory linkage of hypondroplasia to chromosome 4p. *Amer. J. Med. Genet.*
57, 505-506.

Hediger R., Ansari H.A. and Stranzinger G.F. (1991) Chromosome banding and gene localization support extensive conservation of chromosome structure between cattle and sheep. *Cytogenet. Cell Genet.* **57**, 127-134.

Hellerman E.M., Kashi Y., Nave A., Beckmann J.S. and Soller M. (1986) Restriction fragment length polymorphism in dairy cattle and their utilization in genetic improvement. *Rev. Anim. Prod.* **22**, 31-38.

Hetzel D.J.S. (1991) The use of reference families for genome mapping in domestic livestock. In *Gene-mapping techniques and applications.* (ED. Schook L.B, Lewin H.A. and McLaren D.G.) Marcel Dekker Inc. 51-64.

Higgs D.R., Googbuorn S.E.Y., Wainscoat J.S., Clegg J.B. and Weatherall D.J. (1981) Highly variable regions of DNA flank the human globin genes. *Nucl. Acids Res.* **9**, 4213-4224.

Hill A.V.S. and Jeffreys A.J. (1985) Use of minisatellite DNA probes for determination of twin zygosity at birth. *Lancet.* **2**, 1394-1395.

Hines H.C., Kiddy C.A., Brum E.W. and Arave C.W. (1969) Linkage among cattle blood and milk polymorphisms. *Genetics.* **62**, 401-412.

Hines H.C. (1990) Genetic markers for quantitative trait loci in dairy cattle. *Proc. of 4th World Congress on Genetics applied to livestock production. Edinburgh.* **X111**, 121-124.

Hines H.C. and Larsen B. (1990) Evidence against linkage between two pairs of bovine blood group loci: B-Z and F-S. *Anim. Genet.* **21**, 431-432.

Hoeschele I. and Meinert T.R. (1990) Association of genetic defects with yield and type traits: the weaver locus effect on yield. *J. Dairy Sci.* **73**, 2503-2515.

Hohenhorst J., Fries R., Vogeli P. and Stranzinger G. (1994) Use of microsatellites for parentage control in pigs. *Anim. Genet.* **25**, Supplement 2. 33.

Holmes N.G. (1994) Microsatellite markers and the analysis of genetic disease. *Br. vet. J.* **150**, 1-11.

Hughes A.E. and Litt M. (1993) Optimization of microsatellite analysis for genetic mapping. *Genomics.* **15**, 433-434.

Hutt F.B. (1958) Innate disease in *Genetic resistance to disease in domestic animals.* Constable and Company Ltd. 11-32.

Jabs E.W., Li X., Scott A.F., Meyers G., Chen W., Eccles M., Mao J., Charnas L.R., Jackson C.E. and Jaye M. (1994) Jackson-Weiss and Crouzon syndromes are allelic with mutations in febroblast growth factor receptor 2. *Nature Genet.* **8**, 275-279.

Jacob C.O. and Hwang F. (1992) Defenition of microsatellite size variants for TNFA and HSP70 in autoimmune and nonautoimmune mouse strains. *Immunogenetics.* 182-188.

Jamieson A. (1965) The genetics of transferrins in cattle. *Heridity.* **20**, 419-441.

Jamieson A. (1994) The effectiveness of using co-dominant polymorphic allelic series for 1) checking pedigrees and 2) distinguishing full-sib pair members. *Anim. Genet.* **25**, 37-44.

Jeffreys A.J. Wilson V. and Thein S.L. (1985) Individual specific fingerprint of human DNA. *Nature.* **316**, 76-79.

Jeffreys A.J., Wilson V and Thein S.L. (1985a) Hypervariable minisatellite regions in the human DNA. *Nature.* **314**, 67-73.

Jeffreys A.J., Brookfield J.F.Y. and Semeonoff R. (1985b) Positive identification of an immigration test-case using human DNA fingerprints. *Nature.* **317**, 818-819.

Jeffreys A.J., Wilson V., Thein S.L., Weatherall D.J. and Ponder B.A.J. (1986) DNA fingerprints and segregation analysis of multiple markers in human pedigrees. *Amer. J. Hum. Genet.* **39**, 11-24.

Jeffreys A.J. and Morton D.B. (1987) DNA fingerprints of dogs and cats. *Anim. Genet.* **18**, 1-15.

Jeffreys A.J., Wilson V., Neumann R. and Keyte J. (1988) Amplification of human minisatellites by the polymerase chain reaction. *Nucl. Acids. Res.* **16**, 10953-10971.

Jeffreys A.J., Turner M. and Debenham P. (1991a) The efficiency of multilocus DNA fingerprint probes for individualization and establishment of family relationship, determined from extensive case work. *Amer. J. Hum. Genet.* **48**, 824-840.

Jeffreys A.J., Macleod A., Tamaki K., Neil D.L and Mossekton D.G. (1991b) Minisatellite repeat coding as a digital approach to DNA typing. *Nature.* **354**, 204-209.

Jellinek W.R. and Schmid C.W. (1982) Repetitive sequences in eukaryotic DNA and their expression. *Ann. Rev. Biochem.* **51**, 813-844.

Johnson D.E. and Williams L.T. (1993) Structural and functional deversity in the FGF receptor multigene family. *Adv. Cancer Res.* **60**, 1-41.

Jolly R.D. and Blakemore W.F. (1973) Inherited lysosomal storage diseases: An essay in comparative medicine. *Vet. Rec.* **92**, 391-400.

Jones K.L. (1988) *Smith's recognizable patterns of human malformation.* 4th edition. Philadelphia: W.B. Saunders.

Jorgensen C.B., Konfortov B.A., Barendse W. and Miller J.R. (1995) Two polymorphic bovine microsatellite loci. *Anim. Genet.* **25**, 132.

Karl S.A., Bowen B.W. and Avise J.C. (1992) Global population genetic structure and male-mediated gene flow in the green turtle (chelonia mydas): RFLP analysis of anonymous nuclear loci. *Genetics.* **131**, 163-173.

Kashi Y., Soller M., Hallerman E. and Beckmann J.S. (1986) Restriction fragment length polymorphisms in dairy cattle genetic improvement. *Proceedings of the 3rd World Congress on Genetics Applied to Livestock Production. Lincoln, U.S.A.* 57-63.

Kashi Y., Lipkin E., Darvasi A., Nave A., Gruenbaum Y., Beckmann J.S. and Soller M. (1990a) Parentage identification in the bovine using deoxyribonucleic acid fingerprints. *J. Dairy Sci.* **73**, 3306-3311.

Kashi Y., Hallerman E. and Soller M. (1990b) Marker-assisted selection of candidate bulls for progeny testing programmes. *Anim. Prod.* **51**, 63-74.

Katoh O., Hattori Y., Sasaki H., S et al. (1993) Isolation of the complementary DNA encoding a nmouse heparin binding growth factor receptor with the use of a unique kinase insert sequence. *Cancer Res.* **53**, 1136-1141.

Kaukinen J. and Varvio S. (1993) Eight polymorphic bovine microsatellites. *Anim. Genet.* **24,** 148.

Keegan K., Johnson D.E., Williams L.T. and Hayman M.J. (1991) Isolation of an additional member of the fibroblast growth factor receptor family, FGFR-3. *Proc. Natl. Acad. Sci. USA.* **88,** 1095-1099.

Kemp D.L., Coppel R.L. and Anders R.F. (1987) Repetitive proteins and genes of malaria. *Annu. Rev. Microbiol.* **41**, 181-208.

Kemp S.J. and Teale A.J. (1991) Dinucleotide repeat polymorphism at the bovine locus FSHB. *Anim. Genet.* **22**, 435.

Kerem B.S., Rommens J.M., Buchanan J.A., Markiewicz D., Cox T.K., Chakravarti A., Buchwald M. and Tsui L.C. (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science.* **245**, 1073-1080.

Kidd K.K., Osterhoff D., Erhard L. and Stone W.H. (1974) The use of genetic relationships among cattle breeds in the formulation of rational breeding policies: an example with South Devon (South Africa) and Gelbvieh (Germany). *Anim. Blood Groups Biochem. Genet.* **5**, 21-28.

Kidd K.K. (1974) Biochemical polymorphisms, breed relationships, and germplasm resources in domestic cattle. *Proceedings of the 1st World Congress of Genetics Applied to Livestock Production. Madraid.* **1**, 321-328.

Kimpton C.P., Gill P., Walton A., Urquhart A., Millican E.S. and Adams M. (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR methods and applications.* 13-22.

Kimura M. and Crow. (1964) The number of alleles that can be maintained in a finite population. *Genetics.* **49**, 725-738.

Kimura M. (1968) Evolutionary rate at the molecular level. *Nature.* **217**, 624-626.

Kimura M. and Ohta T. (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci. USA.* **75**, 2868-2872.

Klagsbrun M. and Baird A. (1991) A dual receptor system is required for basic fibroblast growth-factor activity. *Cell.* **67**, 229-231.

Kluge R. and Geldermann H. (1982) Effects of marked chromosome sections on quantitative traits in the mouse. *Theor. Appl. Genet.* **62**, 1-4.

Knott T.J., Wallis S.C., Pease R.J., Powell L.M. and Scott J. (1986) A hypervariable region 3' to the human apolipoprotein B gene. *Nucl. Acids Res.* **14**, 9215-9216.

Koorey D.J., Bishop G.A. and Mccaughan G.W. (1993) Allele non-amplification - a source of confusion in linkage studies employing microsatellite polymorphisms. *Hum. Mol. Genet.* **2**, 289-291.

Kwiatkowski D.J., Henske E.P., Weimer K., Ozelius L., Gusella J.F. and Haines J. (1992) Construction of a GT polymorphism map of human 9q. *Genomics.* **12**, 229-240.

Lajeunie E., Ma H.W., Bonaventure J., Munnich A., Le-Merrer M. and Renier D. (1995) FGFR2 mutations in Pfeiffer syndrome. *Nature Genet.* **9**, 108.

Lander E.S. and Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics.* **121**, 185-199.

Lander E.S. (1991) Invited editorial: research on DNA typing catching up with courtroom application. *Amer. J. Hum. Genet.* **48**, 819-823.

Larsen B. (1982) On the bovine blood group system. *Anim. Blood Grp. Biochem. Genet.* **13**, 115-121.

Larsen B., Christensen K. and Agerholm J.S. (1992) A possible location of the bovine Z blood group system on chromosome 8. *Anim. Genet.* **23**, Supplement 1. 77.

La-spada A.R., Wilson E.M., Lubahn D.B., Harding A.E. and Fischbeck K.H. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular dystrophy. *Nature.* **352**, 77-79.

Lathrop G.M., Lalouel J.M., Julier C. and Ott J. (1984) Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA.* **81**, 3443-3446..

Lathrop M., Cartwright P., Wright S., Nakamura Y. and Georges M. (1991) Data analysis for linkage studies. *Gene-Mapping Techniques and Applications.* (ED. Schook L.B, Lewin H.A, McLaren D.G.) 177-197.

Lebowitz R.J., Soller M. and Beckmann J.S. (1987) Trait based analysis for the detection of linkage between marker loci and quantitative trait loci in cross between inbred lines. *Theor. Appl. Genet.* **73**, 556-562.

Lehninger A.L. (1982) Aminoacids and peptides. *Principles of Biochemistry.* Worth Publishers INC. 95-120.

Le-Merrer M., Rousseau F., Legeai-Mallet L., Landais J.C., Pelet A., Bonaventure J., Sanak M., Weissenbach J., Stoll C., Munnich A. and Maroteaux P. (1994) A gene for achondroplasia maps to chromosome 4p. *Nature Genet.* **6**, 318-321.

Leveziel H. and Hines H.C (1984) Linkage in cattle between the major histocompatibility complex (BOLA) and the M blood group system. *Genet. Select. Evol.* **16**, 405-415.

Levinson G. and Gutman G.A. (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**, 203-221.

Lin B., Nasir J., Kalchman M.A., McDonald H., Zeisler J., Goldberg P. and Hayden M.R (1995) Structural analysis of the 5' region of mouse and human huntington disease genes reveals conservation of putative promoter region and di- and trinucleotide polymorphisms. *Genomics.* **25**, 707-715.

Lindberg P.G. and Andersson L. (1988) Close association between DNA polymorphism of bovine major histocompatibility complex class I genes and serological BoLA-A specificities. *Anim. Genet.* **19**, 245-255.

Lipkin E., Anne shalom., Khatib H., Soller M and Friedmann A. (1993) Milk as a source of deoxyribonucleic acid and as a substrate for the polymerase chain reaction. *J. Dairy Sci.* **76,** 2025-2032.

Litt M and Luty J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within cardiac muscle actin gene. *Amer. J. Hum. Genet.* **44,** 397-401.

Litt M., Hauge X. and Sharma V. (1993) Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. *BioTechniques.* **15**, 280-284.

Loftus R.T., MacHugh D.E., Ngere L.O., Balain D.S., Badi A.M., Bradley D.G. and Cunningham E.P. (1994) Mitochondrial genetic variation in European, African and Indian cattle populations. *Anim. Genet.* **25**, 265-271.

Luty J.A., Guo Z., Willard H.F., Ledbetter D.H., Ledbetter S. and Litt M. (1990) Five polymorphic microsatellite VNTRs on the human X chromosome. *Amer. J. Hum. Genet.* **46**, 776-783.

MacHugh D.E., Loftus R.T., Bradley D.G., Sharp P.M and Canningham E.P. (1994) Microsatellite DNA variation within and among European cattle breeds. *Proc. Royal Society of London Sec B.* **256,** 25-31.

MacLennan D.H. and Phillips M.S. (1992) Malignant hyperthermia. *Science.* **256**, 789-794.

Mahadevan M., Tsilfidis C., Sabourin L. et al (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the untranslated region of the gene. *Science.* **255**, 1253-1255.

Maichele A.J. and Chamberlain J.S. (1992) Cross-species conservation of a polymorphic dinucleotide repeat in the dystrophin gene. *Mammalian Genome.* **3**, 290-292.

Mayrand P.E., Corcoran K.P., Ziegle J.S., Robertson J.M., Hoff L.B. and Kronick M.N. (1992) The use of fluorescence detection and internal lane standards to size PCR products automatically. *Appl. Theor. Electrophoresis.* **3**, 1-11.

McClure T.J. (1952) Correlation study of bovine erythrocyte antigen A and butter fat test. *Nature.* **17**, 327.

McCombie W.R., Martin-Gallardo A., Gocayne J.D. et al (1992) Expressed genes, *Alu* repeats and polymorphisms in cosmids sequenced from chromosome 4p16.3. *Nature Genet.* **1,** 348-353.

McDermid E.M., Agar N.S. and Chai C.K. (1975) Electrophoretic variation of red cell enzyme systems in farm animals. *Anim. Blood Grps. Biochem. Genet.* **6**, 127-174.

McKusick V.A., Kelly T.E. and Dorset J.P. (1973) Observations suggesting allelism of the achondroplasia and hypchondroplasia genes. *J. Med. Genet.* **10**, 11-16.

McKusick V.A. (1980) The anatomy of the human genome. *J. Heridity.* **71**, 370-391.

Medjugorac I., Kustermann W., Lazar P., Russ I. and Pirchner F. (1994) Marker-

derived phylogeny of European cattle supports demic expansion of agriculture. *Anim. Genet.* **25**, 19-27.

Megnegneau B., Debets F. and Hoekstra R.F. (1993) Genetic variability and relatedness in the complex group of aspergilli based on random amplification of polymorphic DNA. *Curr. Genet.* **23**, 323-329.

Mellersh C., Holmes N., Binns M. and Sampson J. (1994) Dinucleotide repeat polymorphisms at four canine loci (LEI 003, LEI 007, LEI 008 and LEI 015). *Anim. Genet.* **25**, 125-126.

Mengers A., Schwenger B. and Simon D. (1994) Estimating genetic distance between cattle breeds on the DNA level. *Proceedings of the 24th International Congress on Animal Genetics. Prague, Czech. Rep.*

Miesfeld R., Krystal M. and Arnheim N. (1981) A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human alpha and B globin genes. *Nucl. Acids Res.* **9**, 5931-5947.

Moore S.S., Sargeant L.L., King T.J., Mattick J.S., Georges M and Hetzel D.J.S. (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics.* **10,** 654-660.

Moore S.S., Barendse W., Berger K.T., Armitage S.M and Hetzel D.J.S. (1992) Bovine and ovine DNA microsatellites from the EMBL and GENEBANK databases. *Anim. Genet.* **23**, 463-467.

Moore S.S., Byrne K., Berger K.T., Barendse W., McCarthy F., Womack J.E. and Hetzel D.J.S. (1994) Characterization of 65 bovine microsatellites. *Mammalian Genome.* **5**, 84-90.

Morton N.E. (1955) Sequential tests for the detection of linkage. *Amer. J. Hum. Genet.* **7**, 277-318.

Morral N. and Estivill X. (1992) Multiplex PCR amplification of 3 microsatellites within the CFTR gene. *Genomics.* **13**, 1362-1364.

Moscatelli D. (1987) High and low affinity binding-sites for basic fibroblast growth factor on cultured cells - absence of a role for low affinity binding in the stimulation of plasminogen activator production by bovine capillary endothelial-cells. *J. Cell. Physiol.* **131**, 123-130.

Muenke M., Schell U., Hehr A. et al. (1994) A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nature Genet.* **8**, 269-274.

Muenke M. and Schell U. (1995) Fibroblast growth factor receptor mutations in human skeletal disorders. *Trends Genet.* **11**, 308-313.

Muggli-Cockett N.E and Stone R.T. (1988) Identification of genetic vatiation in the bovine major histocompatibility complex DR like genes using sequenced bovine genomic probes. *Anim. Genet.*. **19,** 213-225.

Murray V., Monchawin C. and England P.R. (1993) The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nucl. Acids Res.* **21**, 2395-2398.

Myers R.M., Larin Z. and Maniatis T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science.* **2**, 1242-1246.

Nadeau J.H and Taylor B.A. (1984) Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc. Natl. Acad. Sci. USA.* **81**, 814-818.

Naik S.N (1978) Origin and domestication of zebu cattle (Bos *indicus*). *J. Hum. Evol.* **7**, 23-30.

Nair P.G. (1957) Studies on association between cellular antigens and butterfat percentage in dairy cattle. *Dairy Sci. Abstr.* **19**, 369.

Nakamura Y., Julier C., Wolff R., Holm T., O'Connell P., Leppert M. and White R. (1987) Characterisation of a human midisatellite sequence. *Nucl. Acids Res.* **15**, 2537-2546.

Nakamura Y., Lathrop M., O'Connel P., Leppert M., Lalouel J.M. and White R. (1988) A primary map of ten DNA markers and two serological markers for human chromosome 19. *Genomics.* **3,** 67-71.

Neimann-Sorensen A. (1956) Blood groups and breed structure as exemplified by three Danish breeds. *Acta. Agric. Scand.* **8,** 115-137.

Neimann-Sorensen A. and Robertson A. (1961) The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta. Agric. Scand.* **11**, 163-196.

Nei M. (1972) Genetic distance between population. *Amer. Naturalist.* **106**, 283-292.

Nei M. and Roychoudhury A.K. (1972) Gene differences between Caucasian, Negro and Japanese populations. *Science.* **177,** 434-436.

Nei M., Tajima F. and Tatwno Y. (1983) Accuracy of estimated phylogenetic trees

from molecular data. *J. Mol. Evol.* **19**, 153-170.

Nei M. (1987) *Molecular Evolutionary Biology.*

Nei M. and Takezaki N. (1994) Estimation of genetic distances and phylogenetic trees from DNA analysis. *Proceedings of the 5th World Congress on Genetics applied to livestock production. Guelph, Ontario, Canada.* **XX1**, 405-412.

Nelson D.L. (1991) From linked marker to disease gene - current approaches. *Strategies for gene mapping.* Ed. Schook L.B., Marcel Dekker Inc. NY.

Nicholas F.W. (1987) Single genes in animal breeding. *Veterinary Genetics.* Oxford University Press. 343-364.

Oberklaid F., Danks D.M., Jensen F., Stace L. and Rosshandler S. (1979) Achondroplasia and hypochondroplasia. *J. Med. Genet.* **16**, 140-146.

O'Brien S.J. (1991) Mammalian genome mapping: lessons and prospects. *Curr. Opin. Genet. Devpt.* **1**, 105-111.

O'Brien S.J., Womack J.E., Lyons L.A., Moore K.J., Jenkins N.A. and Copeland N.G. (1993a) Anchored reference loci for comparative genome mapping in mammals. *Nature Genet.* **3**, 103-112.

O'Brien P.J., Shen H., Cory C.R. and Zhang X. (1993b) Use of a DNA-based test for the mutation associated with porcine stress syndrome (malignant hyperthermia in 10,000 breeding swine. *JAVMA.* **203**, 842-851.

Oldridge M., Wilkie A.O.M., Slaney S.F. et al. (1995) Mutations in the third immunoglobulin domain of the fibroblast growth factor receptor-2 gene in Crouzon syndrome. *Hum. Mol. Genet.* **4**, 1077-1082.

Oliver R.A., Brown P., Spooner R.L., Joosten I and Williams J.L. (1989) The analysis of antigen and DNA polymorphisms within the bovine major histocompatibility complex: 1 the class 1 antigens. *Anim. Genet.*. **20,** 31-41.

Orita M.H., Iwahana H., Kanazawa K., Hayashi K. and Sekiya T. (1990) Detection of polymorphisms of human DNA by gel electrophoresis as single stranded conformation polymorphism. *Proc. Natl. Acad. Sci. USA.* **86**, 2766-

Orkin S.H. (1986) Reverse genetics and human disease. *Cell.* **47**, 845-850.

Ornitz D.M. and Leder P. (1992) Ligand specificity and heparin dependence of fibroblast growth factor receptors 1 & 3. *J. Biol. Chem.* **267**, 16305-16311.

Ornitz D.M., Herr A.B., Nilsson M., Westman J., Svahn J. and Waksman G.

(1995) FGF binding and FGF receptor activation by synthetic heparan-derived di- and trisaccharides. *Science.* **268**, 432-436.

Orr-Urteger A., Bedford M.T., Burakova T., Arman E., Zimmer Y., Yayon A., Givol D. and Lonai P. (1993) Developmental localization of the splicing alternatives of fibroblast growth factor receptor-2 (FGFR2). *Develop. Biol.* **158**, 475-486.

Ott J. (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *Amer. J. Hum. Genet.* **51**, 283-290.

Partanen J., Makela T.P., Eerola E., Korhonen J., Hirvonen H., Claesson-Welsh L. and Alitalo K. (1991) FGFR-4, a novel acidic fibroblast growth factor receptor with a distinct expression pattern. *EMBO.* **10**, 1347-1354.

Pasquale E.B. (1990) A distinctive family of embryonic protein tyrosine kinase receptors. *Proc. Natl. Acad. Sci. USA.* **87**, 5812-5816.

Paterson A.H., Lander E.S., Hewtit J.D., Peterson S. and Lincoln S.E. (1988) Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature.* **335**, 721-726.

Pauli R.M., Conroy M.M., Langer L.O., McLone D.G., Naidich T., Franciosi R., Ratner I.M. and Copps S.C. (1983) Homozygous achondroplasia with survival beyond infancy. *Amer. J. Med. Genet.* **16**, 459-473.

Pauling L., Itano H., Singer A. et al (1949) Sickle cell anaemia, a molecular disease. *Science.* **110**, 543-545.

Payne W.J.A. (1991) Demestication: a step forward in civilisation. *Cattle Genetic Resources.* ed. Hickman. World Animal Series, Elsevier Science Publishers. 51-72.

Pemberton J.M., Slate J., Bancroft D.R. and Barrett J.A. (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Mol. Ecol.* **4**, 249-252.

Pena S.D.J., Chakraborty R., Epplen J.T and Jeffreys A.J. (1993) *DNA fingerprinting: State of the science.* Birkhauser, Basel.

Penedo M.C.T and Medrano J.F. (1992) Genetic analysis of steroid 21-hydroxylase microsatellite DNA (CYP21) polymorphism in four breeds of cattle. *Anim. Genet.* **23,** Suppl.1 55-56.

Penny D., Hendy M.D. and Steel M.A. (1992) Progress with methods for constructing evolutionary trees. *Trends Ecol Evol.* **7**, 73-79.

Pepin L., Amigues Y., Lepingle A., Berthier J., Bensaid A. and Vaiman D. (1995)

Sequence conservation of microsatellites between Bos taurus (cattle), Capra hircus (goat) and related species. Examples of use in parentage testing and phylogeny analysis. *Heridity*. **74**, 53-61.

Perret J., Shia Y.C., Fries R., Vassart G. and Georges M. (1990) A polymorphic satellite sequence maps to the pericentric region of the bovine Y chromosome. *Genomics*. **6**, 482-490.

Peters K., Werner S., Chen G. and Williams L. (1992) 2 FGF receptor genes are differentially expressed in epithelial and mesenchymal tissues during limb formation and organogenesis in the mouse. *Development*. **114**, 233-243.

Peters K., Ornitz D., Werner S. and Williams L. (1993) Unique expression pattern of the FGF receptor 3 gene during mouse organogenesis. *Develop. Biol.* **155**, 423-430.

Peters K., Werner S., Liao X., Wert S., Whitsett J. and Williams L. (1994) Targeted expression of a dominant negative FGF receptor blocks branching morphogenesis and epithelial differentiation of the mouse lung. *EMBO*. **13**, 3296-3301.

Pirchner F. (1983) *Population genetics in animal breeding*. 2nd edn. Plenum press. New York and London.

Prasse K.W. (1977) Pyruvate kinase deficiency anaemia. In *Current Veterinary Therapy. VI Small Animal Practice*. (ED. Kirk R.W) Saunders Philadelphia. 434-435.

Prince J.P. and Tanksley S.D. (1992) Restriction fragment length polymorphism in plant breeding and genetics. *Proc. Royal Society of Edinburgh. Section B*. **99**, 23-29.

Proudfoot N.J., Gill A. and Maniatis T. (1982) The structure of the human zeta-globin gene and a closely linked near identical pseudo gene. *Cell*. **31**, 553-563.

Rapraeger A.C., Krufka A. and Olwin B.B. (1991) Requirement of heparan-sulphate for BFGF-mediated fibroblast growth and myoblast differentiation. *Science*. **252**, 1705-1708.

Reardon W., Winter R.M., Rutland P. Pulleyn L.J., Jones B.M. and Malcolm S. (1994) Mutations in the fibroblast growth factor receptor 2 gene cause Crouzon Syndrome. *Nature Genet*. **8**. 98-103.

Reed P.W., Davis J.L., Copeman J.B. et al. (1994) Chromosome-specific microsatellite sets for fluorescence-based, semiautomated genome mapping. *Nature Genet*. **7**, 390-395.

Reeders S.T., Breuning M.H., Davies K.E., Nicholls R.D., Jarman A.P., Higgs D.R., Pearson P.L. and Weatherall D.J. (1985) A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature.* **317**, 542-544.

Rendel J. (1958) Studies of cattle blood groups. *Acta. Agric. Scand.* **8**, 131-161.

Rendel J. (1961) Relationships between blood groups and fat percentage of the milk in cattle. *Nature.* **189**, 408-409.

Reynolds J.B.S., Weir B.S. and Cockerham C.C. (1983) Estimation of coancestory coefficient: Basis for a short-term genetic distance. *Genetics* **105**, 767-779.

Richardson N.E., Butress N., Feinstein A., Stratil A. and Spooner R.L. (1973) Structural studies on individual components of bovine transferrin. *Biochem. J.* **135**, 87-92.

Richards R.I. and Sutherland G.R. (1992) Heritable unstable DNA sequences. *Nature Genet.* **1**, 7-9.

Rigby P.W.J., Dieckmann M., Rhodes C. and Berg P. (1977) Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**, 237-251.

Riordan J.R., Rommens J.M., Kerem B.S. et al (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science.* **245**, 1066-1073.

Rommens J.M., Iannuzzi M.C., Kerem B. et al (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science.* **245**, 1059-1065.

Ron M., Band M., Yanai A. and Weller J.I. (1994) Mapping quantitative trait loci with DNA microsatellites in a commercial dairy cattle population. *Anim. Genet.* **25**, 259-264.

Ross D. and Larsen B. (1981) Confirmation of F_2 allele in the bovine F blood group system. *Anim. Blood Grps. Biochem. Genet.* **12**, 211-213.

Rousseau F., Bonaventure J., Legeai-Mallet L., Pelet A., Rozet J.M., Maroteaux P., Le-Merrer M. and Munnich A. (1994) Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. *Nature.* **371**, 252-254.

Royle N.J., Clarkson R.E., Wong Z. and Jeffreys A.J. (1988) Clustering of hypervariable minisatellite in the proterminal regions of human autosomes. *Genomics.* **3**, 352-360.

Rudolph J.A., Spier S.J., Byrns G., Rogas C.V., Bernoco D. and Hoffman E.P.

(1992) Periodic paralysis in quarter horses: a sodium channel mutation disseminated by selective breeding. *Nature Genet.* **2**, 144-147.

Rutland P., Pulleyn L.J., Reardon W. et al. (1995) Identical mutations in the FGFR2 gene cause both Pfeiffer and Crouzon syndrome phenotypes. *Nature Genet.* **9**, 173-176.

Rzhetsky A. and Nei M. (1992) A simple method for estimating and testing minimum-evolution trees. *Mol. Biol. Evol.* **9**, 945-967.

Saiki R.K., Scharf S., Faloona F., Mullis K.B., Horn G.T., Erlich H.A. and Arnheim N. (1985) Enzymatic amplification of beta globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science.* **230,** 1350-1354.

Saitou N. and Imanishi T. (1989) Relative efficiencies of the Fitch - Margoliash, maximum parsimony, maximum likelihood, minimum evolution and neighbour joining methods of phylogenetic tree construction in optining the correct tree. *Mol. Biol. Evol.* **6**, 514-525.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) Molecular cloning. A laboratory manual (2nd ed: Nolan C.) Cold Spring Harbor Laborotory Press. **1,2 & 3.**

Sanger F. and Coulson A.R. (1975) A rapid method for determining sequences in DNA by primed sythesis with DNA polymerases. *J. Mol. Biol.* **94**, 441-446.

Sanger F., Nicklen S. and Coulson A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5468.

Schwenger B., Schober S. and Simon D. (1993) DUMPS cattle carry a point mutation in the uridine monophosphate synthase gene. *Genomics.* **16**, 241-244.

Schwengel D.A., Jedicka A.E., Nanthakumara E.J., Weber J.L. and Levitt R.C. (1994) Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. *Genomics.* **22**, 46-54

Shah K., Astley R. and Cameron A. (1973) A thanatophoric dwarfism. *J. Med. Genet.* **10**, 243-252.

Shanks R.D., Dombrowski D.B., Harpestad G.W. and Robinson J.L. (1984) Inheritance of UMP Synthase in dairy cattle. *J. Hered.* **75**, 337-340.

Sheffield V.C., Cox D.R., Lerman L.S. and Myers R.M. (1989) Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA.* **86**, 232-236.

Shiang R., Thompson L.M., Zhu Y.Z. et al. (1994) Mutations in the transmembrane domain of FGFR3 cause the most common form of genetic dwarfism, achondroplasia. *Cell.* **78**, 335-342.

Shuster D.E., Kehrli M.E., Ackermann M.R. and Gilbert R.O. (1992) Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc. Natl. Acad. Sci. USA.* **89**, 9225-9229.

Silver H. (1989) Paternity testing. *Cr. Rev. Clin. Lab. Sci.* **27**, 391-408.

Skolnick M.H. Wallace R.B. (1988) Simultaneous analysis of multiple polymorphic loci using amplified sequence polymorphisms (ASPs). *Genomics.* **2**, 273-279.

Slightom J.L., Blechl A.E. and Smithies O. (1980) Human fetal γ and α globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell.* **21**, 627-638.

Smeets H.J.M., Brunner H.G., Ropers H.H. and Wierinnga B. (1989) Use of variable simple sequence motifs as genetic markers: application of study of myotonic dystrophy. *Hum. Genet.* **83**, 245-251.

Smith C.A.B. (1957) Counting methods in genetic statistics. *Ann. Hum. Genet.* **21**, 254-274.

Smith C.A.B. (1977) A note on genetic distance. *Ann. Hum. Genet.* **40**, 463-479.

Smith C. and Simpson S.P. (1986) The use of genetic polymorphisms in livestock improvement. *J. Anim. Breed. Genet.* **103**, 205-217.

Sneath P.H.A. and Sokal A.A. (1973) The principles and practice of numerical classification. *Numerical Taxonomy.* Sanfrancisco: W.H.Freeman.

Solinas-Toldo S., Lengauer C. and Fries R. (1995) Comparative genome map of human and cattle. *Genomics.* **27**, 489-496.

Soller M. (1978) The use of loci associated with quantitative effects in dairy cattle improvement. *Anim. Prod.* **27**, 133-139.

Soller M. and Beckmann J.S. (1982) Restriction fragment length polymorphism and animal improvement. *Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production. Madraid.* **V1**, 396-404.

Soller M. and Beckmann J.S. (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theor. Appl. Genet.* **67**, 25-33.

Soller M. and Beckmann J.S. (1985) Restriction fragment length polymorphisms

and animal genetic improvement. *Rev. Rural Sci.* **6**, 10-16.

Soller M. and Beckmann J.S. (1990) Molecular mapping of quantitative genes. *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production. Edinburgh.* **XIII**, 93-96.

Soller M. (1990) Genetic mapping of the bovine genome using deoxyribonucleic acid-level markers to identify loci affecting quantitative traits of economic importance. *J. Dairy Sci.* **73**, 2629-2646.

Southern E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Spooner R.L. (1967) Blood groups in animals and their practical application. *Vet. Rec.* 699-708.

Spooner R.L. (1980) Blood groups in animals. In *Scientific foundations of veterinary medicine.* Ed. Phillipson A.T., Hall L.W., Pritchard W.R. William Heinemann Medical Books Ltd (1st edition).

Stallings R.L., Ford A.F., Nelson D., Torney D.C., Hildrbrand C.E. and Moyzist R.K. (1991) Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. *Genomics.* **10**, 807-815.

Stallings R.L. (1994) Distribution of trinucleotide microsatellites in different categories of mammalian genomic sequence: implications for human genetic disease. *Genomics.* **21**, 116-121.

Stanescu R. Stanescu V. and Maroteaux P. (1990) Homozygous achondroplasia: morphologic and biochemical study of cartilage. *Amer. J. Med. Genet.* **37**, 412-421.

Stark K.L., Mcmahon J.A. and Mcmahon A.P. (1991) FGFR-4, a new member of the fibroblast growth-factor receptor family, expressed in the definitive endoderm and skeletal muscle lineages of the mouse. *Development.* **113**, 641-651.

Steffan P., Eggan A., Dietz A.B., Womack J.E., Stranzinger G. and Fries R. (1993) Isolation and mapping of polymorphic microsatellites in cattle. *Anim. Genet.*. **24,** 121-124.

Stone W.H. (1962) The J substance of cattle. *Ann. N. Y. Acad. Sci.* **97**, 269.

Stoneking M. (1993) DNA and recent human evolution. In *Evolutionary Anthropology.* **2**, 60-73.

Stormont C. (1958) On the applications of blood groups in animal breeding. *Proceedings of the Xth International congress of genetics.* 206-223.

Stormont C. (1967) Contribution of blood typing to dairy science progress (1967) *J. Dairy Sci.* **50,** 253-260.

Stormont C. (1978) The early history of cattle blood groups. *Immunogenetics.* **6,** 1-15.

Stratil A. and Spooner R.L. (1971) Isolation and properties of individual components of cattle transferrin. The role of sialic acid. *Biochem. Genet.* **5**, 347-365.

Stuber C.W., Mull R.H., Goodman M.M., Shaffer H.E. and Weir B.S. (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (Zea mays). *Genetics.* **95**, 225-236.

Suber M.L., Pittler S.J., Qin N. et al (1993) Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta subunit gene. *Proc. Natl. Acad. Sci. USA.* **90**, 3968-3972.

Sullivan K.M., Pope S., Gill P. and Robertson J.M. (1992) Automated DNA profiling by fluorescent labelling of PCR products. *PCR methods and applications.* 34-40.

Superti-Furga A., Eich G., Bucher H.U., Wisser J., Giedion A. Gitzelmann R. and Steinmann B. (1995) A glycine 375 to cysteine substitution in the transmembrane domain of the fibroblast growth factor receptor 3 in a new born with achondroplasia. *Eur. J. Paediatr.* **154**, 215-219.

Suzuki R., Kemp S.J. and Teale A.J (1993) Polymerase chain reaction analysis of mitochondrial DNA polymorphism in N'Dama and Zebu cattle. *Anim. Genet.* **24**, 339-343.

Swofford D.L and Olsen G.J. (1990) Phylogeny reconstruction. In *Molecular Systematics.* ed. Hillis D.M. and Monitz C. Sinaver Associates, Sunderland, USA.

Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl. Acids Res.* **17**, 6463-6471.

Tavormina P.L., Shiang R., Thompson L.M. et al. (1995) Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nature Genetics.* **9**, 321-328.

Teale A.J., Tan S.G. and Tan J.H. (1994) Applications of molecular genetic and reproductive technologies in the conservation of domestic diversity. *Proceedings of the 5th World Congress on Genetics applied to livestock production. Guelph, Ontario, Canada.* **XX1**, 493-500.

Threadgill D.W. and Womack J.E. (1990a) Syntenic conservation between human and cattle. 1.human chromosome 9. *Genomics.* **8**, 22-28.

Threadgill D.W. and Womack J.E. (1990b) Syntenic conservation between human and cattle. II.human chromosome 12. *Genomics.* **8**, 29-34.

Threadgill D.W., Krause J.P., Krawetz S.A. and Womack J.E. (1991) Evidence for the evolutionary origin of human chromosome 21 from comparative gene mapping in bovine. *Proc. Natl. Acad. Sci. USA.* **88**, 154-158.

Thomas K.A. (1987) Fibroblast growth-factors. *FASEB J.* **1**, 434-440.

Thompson L.M., Plummer S., Schalling M., Altherr M.R., Gusella J.F., Housman D.E. and Wasmuth. (1991) A gene encoding a fibroblast growth factor receptor isolated from the Huntington disease gene region of human chromosome 4. *Genomics.* **11**, 1133-1142.

Trommelen G.J.M., Daas N.H.G.D., Vijg J. and Uitterlinden A.G. (1993) Identity and paternity testing of cattle: application of a deoxyribonucleic acid profiling protocol. *J. Dairy Sci.* **76**, 1403-1411.

Usha A.P., Simpson S.P. and Williams J.L. (1995) Probability of random sire exclusion using microsatellite markers for parentage verification. *Anim. Genet.* **26**, 155-161.

Vaiman D., Mercier D., Moazami-Goudarzi K., Eaggem A., Ciampolini R., Lepingle A., Velmala R., Kaukinen J., Varvio S.L., Martin P., Leveziel H. and Guerin G. (1994) A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mammalian Genome.* **5**, 288-297.

Valdes A.M., Slatkin M. and Freimer N.B. (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics.* **133**, 737-749.

Vankan D.M., Moore S.S., Bell K. and Hetzel D.J.S. (1994) Evaluation of DNA microsatellite markers for parentage and paternity testing of cattle. *Proceedings of the 24th International Conference on Animal Genetics. Prague, Czech Republic.* **25**, Supplement 2. 41.

Van Vleck L.D (1970) Misidentification and sire evaluation. *J. Dairy Sci.* **53**, 1697-1702.

Vassart G., Georges M., Monsieur R., Brocas H., Lequarre A.S. and Christophe D. (1987) A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science.* **235**, 683-685.

Velinov M., Slaugenhaupt S.A., Stoilov I., Scott C.I., Gusella J.F. and Tsipouras

P. (1994) The gene for achondroplasia maps to the telomeric region of chromosome 4p. *Nature Genet.* **6**, 314-317.

Vergnaud G. (1989) Polymers of random short oligonucleotide detect polymorphic loci in the human genome. *Nucl. Acids Res.* **17**, 7623-7630.

Visscher P.M. and Haley C.S. (1995) Utilizing genetic markers in pig breeding programmes. *Anim. Breed. Absts.* **63**, 1-8.

Wainwright B.J., Scambler P.J., Schmidtke J., Watson E.A., Law H.Y., Farrall M., Cooke H.J., Eiberg H. and Williamson R. (1985) Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature.* **318**, 384-385.

Walker B.A., Murdoch J.L., McKusick V.A., Langer L.O. and Beals R.K. (1971) Hypochondroplasia. *Amer. J. Dis. Child.* **122**, 95-104.

Warner J.P., Barron L.H. and Brock J.H. (1993) A new polymerase chain reaction (PCR) asssay for the trinucleotide repeat that is unstable and expanded on huntington's disease chromosomes. *Mol. Cell. Probes.* **7**, 235-239.

Watanabe T., Hayashi Y., Semba R. and Ogasawara N. (1985) Bovine mitochondrial DNA polymorphism in restriction endonuclease cleavage patterns and the location of the polymorphic sites. *Biochem. Genet.* **23**, 947-957.

Watanabe T., Masangkay J.S., Wakana S., Saitou N. and Tomita T. (1989) Mitochondrial DNA polymorphism in native Philippine cattle based on restriction endonuclease cleavage patterns. *Biochem. Genet.* **27**, 431-438.

Watson J.D., Gilmon M., Witkowski J. and Zoller M. (1992) *Recombinant DNA.* 2nd edition. Scientific American Books. W.H.Freeman & Co. New York. 135-152.

Weaver A.D. (1975) Dwarfism in cattle. In *The Veterinary Annual.* (Ed: Grunsell C.S.G. Hill F.W.G. Wright-scientechnica Bristol **XV**, 7-9.

Weber J.L. (1988) Molecular biology of malaria parasites. *Exp. Parasitol.* **66**, 143-170.

Weber J.L. (1989) Length polymorphism in (dC-dA)n-(dG-dT)n sequences detected using the polymerase chain reaction. *The Polymerase Chain Reaction. Current Comments in Molecular Biology.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 141-146.

Weber J.L and May P.E. (1989) Abundant class of DNA polymorphisms which can be typed using the polymerase chain reaction. *Amer. J. Hum. Genet.* **44**, 388-396.

Weber J.L. (1990a) Human DNA polymorphisms based on length variations in

simple sequence tandem repeats. *Genome Analysis.* **1**, 159-181.

Weber J.L. (1990b) Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics.* **7**, 524-530.

Weber J.L. and Wong C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**, 1123-1128.

Weir B.S. (1990) *Genetic data analysis.* Sinauer Associates Inc. Southerland. MA, USA. 187.

Weissenbach J. (1993) Microsatellite polymorphisms and the genetic linkage map of the human genome. *Curr. Opin. Genet. Devpt.* **3**, 414-417.

Weller J.I., Kashi Y. and Soller M. (1990) Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy Sci.* **73**, 2525-2537.

Weller J.I. and Fernando R.L. (1991) Strategies for the improvement of animal production using marker-assisted selection. *Gene-Mapping Techniques and Applications.* (ed. by Schook L.B, Lewin H.A, McLaren DG.) 305-328.

Wenk R.E., Chiafari F.A., Brooks M.A. and Houtz T.D. (1992) Technical progress in parentage analysis. *Rep. Med.* **12**, 621-642.

White R.L., Leppert M., Bishop T., Barker D. and Berkowitz J. (1985) Construction of linkage maps with DNA markers for human chromosomes. *Nature.* **313**, 101-105.

White R. and Lalouel J.M. (1988) Chromosome mapping with DNA markers. *Sci. Amer.* **258**, 20-29.

Wilkie A.O.M., Slaney S.F., Olridge M. et al. (1995) Apert syndrome results from localized mutations of FGFR2 and is allelec with Crouzon syndrome. *Nature Genet.* **9**, 165-171.

Williams A.F. and Barkclay N.A. (1988) The immunoglobulin superfamily - domains for cell-surface recognition. *Ann. Rev. Immunol.* **6**, 381-408.

Womack J.E. and Moll Y.D. (1986) Gene map of the cow: conservation of linkage with mouse and man. *J. Hered.* **77**, 2-7.

Womack J.E. (1987) Genetic engineering in agriculture: animal genetics and development. *Trends Genet.* **3**, 65-68.

Womack J.E., Threadgell D.W., Moll Y.D., et al (1989) Syntenic mapping of 37

loci in cattle. Chromosomal conservation with mouse and man. *Cytogenet. Cell Genet.* **51**, 1109.

Womack J.E. (1990) Gene mapping in the cow. *Adv. Vet. Sci. Comp. Med.* **34**, 251-271.

Womack J.E., Bolch S.N. and Fries R. (1992) Gene map of the cow (Bos *taurus*). In *Genetic maps: Locus maps of complex genomes.* (ed. O'Brien S.J. 1993 6th edn.) Cold Spring Harbor Laboratory Press. **4**, 264-275.

Womack J.E. (1993) The goals and status of the bovine gene map. *J. Dairy Sci.* **76**, 1199-1203.

Wong Z., Wilson V., Jeffreys A.J. and Thein S.N. (1986) Cloning a selected fragment from a human DNA fingerprint: Isolation of an extremely polymorphic minisatellite. *Nucl. Acids Res.* **14**, 4605-4616.

Wong Z., Wilson V., Patel I., Povey S. and Jeffreys A.J. (1987) Characterisation of a panel of highly variable minisatellite cloned from human DNA. *Ann. Hum. Genet.* **51**, 269-288.

Wyman A.R. and White R. (1980) A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA.* **77**, 6754-6758.

Yayon A., Klagsbrun M., Esko J.D., Leder P. and Ornitz D.M. (1991) Cell-surface, heparin-like molecules are required for binding of basic fibroblast growth-factor to its high-affinity receptor. *Cell.* **64**, 841-848.

Youil R., Kemper B.W. and Cotton R.G.H. (1995) Screening for mutations by enzyme mismatch cleavage with T4 endouclease V11. *Proc. Natl. Acad. Sci. USA.* **92**, 87-91.

Zhang H., Scholl R., Browse J. and Somerville C. (1988) Double stranded DNA sequencing as a choice for DNA sequencing. *Nucl. Acids Res.* **16**, 1220.

Zhang N., Threadgill D.W. and Womack J.E. (1992) Synteny mapping in the bovine: genes from human chromosome 4. *Genomics.* **14,** 131-136.

Ziegle J.S., Su Y., Corcoran K.P., Nie L., Mayrand P.E., Hoff L.B., McBride L.J., Kronick M.N. and Diehl S.R. (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics.* **14**, 1026-1031.

Zneimer S.M. and Womack J.E. (1989) Regional localisation of the fibronectin and gamma crystallin genes to bovine chromosome 8. *Genomics.* **5**, 215-220.

Zuliani G. and Hobbs H.H. (1990) A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. *Amer. J. Hum. Genet.* **46**, 963-969.

List of Publications from this thesis.

1. Usha A.P, Simpson S.P. and Williams J.L (1994) Evaluation of microsatellite markers for parentage verification in cattle. Presented at Proc. of the 24th Conference of International Society of Animal Genetics, Prague. *Animal Genetics, **25***, Supplement 2, 41.

2. Simpson S.P., Usha A.P. and Williams J.L (1994) Characterisation of microsatellite markers for gene mapping and their use in parentage testing. *Proc. of the 45th Annual meeting of the European Association for Animal Production Edinburgh.* **1**, 84.

3. Usha A.P, Simpson S.P. and Williams J.L (1995) Probability of Random sire exclusion using microsatellite markers for parentage verification. *Animal genetics* **26**, 155-161.

4. Usha A.P., Lester D.H., Russel G.C., and Williams.J.L (1995) Excluding mutations in the transmembrane region of FGFR3 as the cause of dominantly inherited Dexter dwarfism (achondroplasia) in cattle. To be presented at the "Mammalian Genetics and Development Workshop" - 20-22nd Nov 1995, London.

5. Williams J.L, Usha A.P., Urquhart B.G. and Kilroy M. (1995) Verifying the identity of bovine semen using DNA microsatellite markers. Submitted to Vet.Rec.

6. Simpson S.P., Usha A.P and Williams J.L (1995) Paternity verification in populations which may be in non-Hardy Weinberg Equilibrium with markers in linkage disequilibrium. Submitted to Genet.Res.

EVALUATION OF MICROSATELLITE MARKERS PARENTAGE VERIFICATION IN CATTLE

A.P.USHA, S.P.SIMPSON AND J.L.WILLIAMS

THE ROSLIN INSTITUTE, MIDLOTHIAN, SCOTLAND UK

Introduction:

(DNA) profiling is an efficient tool for paternity verification and forensic medicine in man (Jeffreys et al 1985). Microsatellite based markers obviate several difficulties associated with other types of markers: they have higher heterozygosities than Restriction Fragment Length Polymorphisms and are more easily interpreted than the patterns generated by minisatellites. Moreover their short repeat length (~100 nucleotides) permit amplification by polymerase chain reaction (PCR) (Saiki et al 1985).

Microsatellite loci have been identified in the bovine genome (Barendse et al 1994, Bishop et al 1994). The genes coding for the bovine MHC class II antigen (DRB3), steroid 21 hydroxylase (CYP21), ß subunit of the follicle stimulating hormone (FSHß) and anonymous clones isolated from plasmid libraries e.g. ETH131 and HEL6 contain highly polymorphic microsatellite sequences.

We have evaluated the usefulness of these five microsatellite based markers for parentage determination when used either independantly or together, and assessed their utility in different breeds of cattle in the UK using appropriate statistical methods.

Materials and Methods:

Twenty individuals in 10 breeds and 15 in another four breeds were chosen such that there were no grandparents in common. Microsatellite loci were amplified by the PCR.

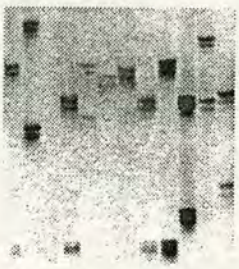
Table 1 Primer sequences used for amplifying the microsatellite loci

1.DRB3	LA 53	CCAGAGTGAGTGAAGTATCTCT
	LA 54	GAGAGTTTCACTGTGCAG
2.CYP21	R	GCTCGCGATCCAACTCCTCCTGAAG
	L	GCCCTGAAGAAGAGAATGGCAACC
3.ETH131	R	GCTGTGATGGTCTACGAATGA
	L	GTGGACTATAGACCATAAGGTC
4.FSHß	R	CAGTTTCTAAGGCTACATGGT
	L	TGGGATATAGACTTAGTGGC
5.HEL6	R	GGACACGACTGAGCAAGTAA
	L	AGGCAGATACATTACCACTA

To estimate the value of the loci for parentage verification, allele frequencies, polymorphic information content(PIC), heterozygosity and PRASE (Probability of Random Sire Excluded) were calculated. PRASE determines the proportion of sires in the population which could be excluded from paternity for a randomly selected dam-progeny pair.

Results:

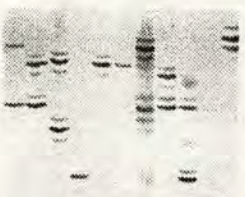
All the loci were found to be highly polymorphic in the breeds studied.



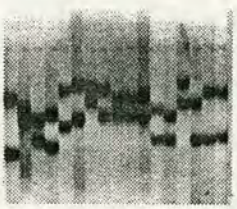
1.DRB3

2.CYP21

3.HEL6



4.ETH131



5.FSHβ

Autoradiograghs of 6% denaturing gels with PCR amplified fragments from the different microsatellite loci.

PRASE values for each breed with the markers used either independantly or together are:

Table 3. Probability of Random Sire Excluded(PRaSE)

Breed	DRB3	CYP21	ETH131	HEL6	FSHβ	Total
AA	0.773	0.688	0.711	0.660	0.526	0.996
AY	0.751	0.671	0.585	0.788	0.718	0.997
BB	0.820	0.730	0.804	0.416	0.595	0.997
BA	0.668	0.685	0.649	0.641	0.700	0.996
CH	0.774	0.621	0.715	0.695	0.490	0.996
DX	0.611	0.409	0.574	0.595	0.775	0.992
FR	0.831	0.812	0.661	0.753	0.810	0.999
GU	0.679	0.722	0.561	0.696	0.509	0.993
HF	0.441	0.748	0.775	0.814	0.628	0.997
JE	0.797	0.756	0.609	0.475	0.829	0.998
LN	0.852	0.586	0.765	0.612	0.730	0.998
SG	0.695	0.775	0.626	0.694	0.808	0.998
BG	0.720	0.582	0.978	0.146	0.405	0.978
SD	0.579	0.752	0.992	0.614	0.536	0.992
Overall	0.720	0.690	0.662	0.623	0.650	0.996

The usefulness of the markers differed between breeds. DRB-3 and CYP-21 being linked(1.4cM), the data for these markers were analysed together. An average probability of exclusion of 0.88 was obtained using both markers. When taken together the results from the five markers gave a 99% probability of identifying an incorrect sire, in all of the 14 breeds.

Table 2 The number of alleles, PIC and heterozygosity values for each breed against marker are given as follows.

Breed	No:of Animals	DRB-3			Microsatellite Loci CYP-21			ETH131			FSHb			HEL6	
		Alleles	PIC	Hetero	Alleles	PIC	Hetero	Alleles	PIC	Hetero	Alleles	PIC	Hetero	Alleles	PIC	Hetero
Aberdeen Angus	20	12	0.86	0.88	10	0.82	0.84	8	0.76	0.4	6	0.65	0.55	9	0.73	0.55
Ayrshire	20	9	0.83	0.85	10	0.83	0.82	9	0.75	0.75	11	0.78	0.45	12	0.84	0.5
Belgian Blue	20	11	0.87	0.65	10	0.83	0.85	15	0.88	0.7	11	0.78	0.79	6	0.63	0.68
Blonde Aquitaine	15	8	0.68	0.27	10	0.83	0.84	10	0.8	0.82	8	0.76	0.47	9	0.76	0.79
Charlaois	20	11	0.87	0.88	10	0.77	0.75	7	0.75	0.78	9	0.83	0.47	5	0.68	0.53
Dexter	15	7	0.73	0.77	6	0.69	0.73	10	0.83	0.7	6	0.68	0.72	13	0.75	0.45
Friesian	20	11	0.87	0.45	13	0.89	0.75	12	0.79	0.81	15	0.9	0.91	11	0.83	0.65
Guernsey	20	11	0.82	0.84	9	0.82	0.84	10	0.73	0.77	13	0.68	0.7	10	0.8	0.82
Hereford	20	9	0.62	0.62	10	0.82	0.7	11	0.88	0.89	10	0.75	0.77	11	0.88	0.6
Jersey	20	11	0.85	0.5	10	0.85	0.7	6	0.72	0.65	11	0.87	0.45	7	0.65	0.65
Limousin	20	14	0.9	0.6	12	0.76	0.77	10	0.85	0.7	11	0.82	0.6	10	0.75	0.78
Simmental	20	11	0.75	0.5	10	0.86	0.7	10	0.77	0.8	11	0.86	0.5	8	0.78	0.65
B.Galloway	15	10	0.86	0.87	7	0.76	0.8	8	0.79	0.93	4	0.51	0.33	2	0.16	0.18
South Devon	15.	7	0.77	0.8	10	0.85	0.8	6	0.75	0.78	7	0.76	1	7	0.66	0.33
Overall	260	23	0.92	0.64	19	0.91	0.78	24	0.89	0.74	21	0.91	0.61	19	0.88	0.57

Conclusions:

1. Microsatellite markers can be efficiently used for parentage verification in cattle.

2. Use of linked markers like DRB-3 and CYP-21 does not reduce PRASE considerably when compared with similar unlinked markers.

3. Use of five markers with average of ten alleles would identify an incorrect parent in 99% of cases in common cattle breeds.

References:

Barendse *et al* (1994). A linkage map of the bovine genome. *Nature Genetics* **6**, 227-235.

Bishop *et al* (1994). A genetic linkage map for cattle. *Genetics* **136**, 619-639.

Jeffreys *et al* (1985). Hypervariable minisatellite regions in the human DNA. *Nature* **354**, 204-209.

Saiki *et al* (1985). Enzymatic amplification of beta globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Nature* **230**, 1350-1354.

CHARACTERISATION OF MICROSATELLITE MARKERS FOR GENE MAPPING AND THEIR USE IN PARENTAGE TESTING

Simpson. S.P., Usha. A.P. and Williams. J.L

(Presented at the 45th Annual meeting of EAAP, Edinburgh. Abstract No: G1.17)

Marker heterozygosity, H, and polymorphism information content, PIC, are frequently used to evaluate new markers for gene mapping and parentage testing. Since new markers are evaluated on a limited number of animals, the precision of H and PIC may be low. We have estimated standard errors for H and PIC for real and simulated data using jack-knifing, i.e. by dropping animals from the data set and recalculating H and PIC. Confidence bands cannot be formed as, due to the loss of alleles from the sample, there is no guarantee that H and PIC lie within the range of jack-knifed estimates. The spread of values can be used, however, to gain insight into the precision of H and PIC. Conventional tests of Hardy-Weinberg and linkage equilibrium based on chi-squared goodness-of-fit tests are often inappropriate since the expected numbers with each genotype can be very small. Bootstrapping, or Monte Carlo resampling, can be used to determine significance levels. The sampling error due to resampling can be reduced by further resampling. The error due to sampling from the population may be very large and can be estimated by jack-knifing. In our studies of linked microsatellite markers we observed large deviations from Hardy-Weinberg and linkage equilibrium. The probability that a random sire can be excluded from parentage measures how useful markers are for parentage testing. Conventional methods assume the population is in equilibrium. We have extended these to account for deviations from equilibrium.

Animal Genetics,
1995, **26**, 155–161

Probability of random sire exclusion using microsatellite markers for parentage verification

A P Usha, S P Simpson, J L Williams

Summary

Many microsatellite sequences have been described in the bovine genome. Being highly polymorphic these have been suggested as markers for parentage verification and individual identification in cattle. We have evaluated the use of five highly polymorphic microsatellite markers for parentage verification in 14 breeds of cattle in the UK. Three of the microsatellite loci occur within introns in genes: *BoLA DRB3*, steroid 21-hydroxylase, and the beta subunit of the follicle-stimulating hormone. The other two are anonymous sites *ETH131* and *HEL6*. Results were analysed by a statistical approach that takes in to account deviations from Hardy–Wienberg equilibrium and linkage disequilibrium for multiple loci. The method of determining the probability of random sire exclusion uses observed genotype frequencies instead of allele frequencies. Independently, the markers used have a probability of between 0·72 and 0·62 of identifying a parentage error, while used together the five markers give, on average across breeds, a probability of 0·99 of excluding an incorrect sire.

Keywords: cattle, microsatellite, parentage

Introduction

DNA profiling has become a major tool in paternity verification and forensic medicine in humans (Jeffreys *et al.* 1985). Several types of markers have been used for creating the profile and each has different advantages and disadvantages associated with it: restriction fragment length polymorphisms (RFLPs) generally suffer from low heterozygosities and low polymorphic information content (PIC) (Botstein *et al.* 1980), while the genetic fingerprints produced by minisatellite probes are difficult to interpret or reproduce owing to the extreme genetic variability revealed (Jeffreys *et al.* 1991).

Microsatellite-based markers overcome many of the difficulties associated with the other marker types. They generally have higher heterozygosities than RFLPs and, in arising from defined loci, results are more easily interpreted than patterns generated by minisatellites. Microsatellite sequences are stretches of tandemly repeated short sequence motifs, one to six nucleotides in length (Litt & Luty 1989; Weber & May 1989). Polymorphism arises from differences in the number of times the motif is repeated. An additional advantage of the microsatellite markers is that, in being 100–300 nucleotides long, they can be amplified by polymerase chain reaction (PCR) (Saiki *et al.* 1985, 1988). There has therefore been interest in using microsatellite markers for paternity verification and individual identification in cattle.

Many microsatellite loci have been described in the bovine genome (Barendse *et al.* 1994; Bishop *et al.* 1994), and about 400 have been characterized (Hetzel 1993). Microsatellite sequences are found in introns in the genes coding for the bovine major histocompatibility complex class II antigen (*DRB3*), (Muggli-Cockett & Stone 1988), steroid 21-hyroxylase (*CYP21*), (Chung *et al.* 1986) and the beta subunit of the follicle stimulating hormone (*FSHB*) (Kim *et al.* 1988). In addition to these the anonymous sequences, *D21S4* (*ETH131*) and *HEL6* are used in this study. All of these microsatellite regions have been shown to be highly polymorphic in cattle (Fries *et al.* 1990; Moore *et al.* 1992; Ellegren *et al.* 1993). In cattle, *DRB3* and *CYP21* loci are present on chromosomes 23 (Fries *et al.* 1986), have been shown to share synteny (Skow *et al.* 1988), are linked with a lod score of 10·4 at a recombination fraction of 0·17 (Creighton *et al.* 1992), and are separated by only 1·4 cM on the latest map of the bovine genome (W. Barendse, personal communication). The other markers are located on different chromosomes and are unlinked.

Techniques for parentage verification in humans have assumed Hardy–Weinberg (HW) equilibrium (Devlin *et al.* 1990), which is unlikely to occur in cattle populations. The statistical tests used to date in livestock have been inappropriate for use with small sample sizes (e.g. Mejdell *et al.* 1993). It is particularly important to account for sampling error when the number of alleles present at a locus is large and

A P Usha
S P Simpson
J L Williams
Roslin Institute,
Edinburgh Research
Station, Roslin,
Midlothian EH25 9PS,
UK

Correspondence: J L Williams.
Accepted 17 January 1995

the number of individuals examined is small. In this study we examine the value of the five microsatellite loci: *DRB3*, *CYP21 FSHB*, *ETH131* and *HEL6* in parentage verification, when used either independently or together, and assess their utility in several breeds of cattle, taking into account deviations from HW equilibrium and using appropriate statistical methods for the small sample sizes.

Materials and methods

Animals

A total of 260 animals in 14 breeds was used in this study, 20 individuals in 10 of the breeds and 15 in the other 4. Samples were provided by the Cattle Blood Typing Service, Roslin, Edinburgh, UK. The pedigrees of the individuals were inspected and those chosen had no common grandparents. The breeds used are listed in Table 2.

DNA preparation

DNA was extracted from peripheral blood lymphocytes as described by Oliver *et al.* (1989) and stored at 4°C.

Analysis of microsatellite length polymorphism

Regions containing the microsatellite repeats were amplified by PCR. The primers used for amplification and the PCR conditions optimized for each locus are given in Table 1.

The PCR was in 20 µl volumes containing: 200 ng genomic DNA; 20 pmoles each oligonucleotide primer; 200 µM each dCTP, dGTP, dTTP; 40 µM dATP; 2·5 µCi ^{35}S dATP (Amersham, Bucks, UK); 10 mM Tris (pH 8·3); 50 mM KCl; 0·001% gelatin; $MgCl_2$; and 0·5 units of *Taq* DNA polymerase (Boerhinger, Mannheim, Germany). The PCR protocol was 94°C denaturation (1 min) followed by 60°C annealing and extension (5 min) for 30 cycles and the products were stored at 4°C. Samples were mixed with formamide loading buffer (50% deionized formamide containing 0·1% w/v bromophenol blue and xylene cyanol), denatured at 94°C (4 min) and then loaded on to 6% denaturing polyacrylamide gel. Electrophoresis was at 40 W for 3 h after which the gels were fixed (10% methanol, 10% acetic acid), dried, and exposed to X-ray film for two days. Apparent sizes of alleles were assigned by comparison with a sequencing ladder from M13.

Table 1. Characterization of bovine microsatellites

Locus	Primer sequence	Annealing temperature (°C)	Mg^{2+} conc.
DRB3	LA53 CCAGAGTGAGTGAAGTATCTCT	60	1·25 mM
	LA54 GAGAGTTTCACTGTGCAG		
CYP21	R GCTCGCGATCCAACTCCTCCTGAAG	60	1·25 mM
	L GCCCTGAAGAAGAGAATGGCAACC		
HEL6	R GGACACGACTGAGCAAGTAA	60	1·5 mM
	L AGGCAGATACATTACCACTA		
FSHB	R CAGTTTCTAAGGCTACATGGT	60	1·25 mM
	L TGGGATATAGACTTAGTGGC		
ETH131	R GCTGTGATGGTCTACGAATGA	55	1·25 mM
	L GTGGACTATAGACCATAAGGTC		

Statistical analysis

To estimate the value of the loci for parentage verification, allele frequencies, PIC, heterozygosity and the probability that a random sire can be excluded (PRASE) were calculated. Monte Carlo techniques were used to test for HW equilibrium, differences in gene frequencies between breeds, and linkage equilibrium (Devlin & Risch 1992; Guo & Thompson 1992).

PIC was calculated as:

$$1-\left(\sum_{i=1}^{n} p_i^2\right)\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_i^2p_j^2$$

where p_i is the frequency of the i^{th} allele (Botstein *et al.* 1980).

Heterozygosity was calculated as:

$$1-\Sigma p_i^2$$

(Ott 1992).

PRASE is the proportion of sires in a population that can be excluded from paternity for a randomly selected dam–calf pair. For populations in HW equilibrium and a single locus, the method of Chakroborty *et al.* (1988) can be used. We extended their methods to account for deviations from HW equilibrium and for analysis of more than one locus, whether in linkage equilibrium or not, by using observed genotypes instead of allele frequencies. In its most general form, calculation of PRASE can be expressed as:

$$\text{PRASE} = \sum_d \sum_s P(G_d)\,P(G_s) \sum_o P(G_o/G_d, G_s)$$
$$\sum_p P(G_p)\,P(\text{excluded}/G_d, G_p, G_o)$$

where $P(G_d)$, $P(G_s)$, $P(G_p)$ are the genotype frequencies of the dams, true sires and putative sires, and dependent on whether or not the population is in HW equilibrium and for more than one locus, and whether or not these are in linkage equilibrium. The term $P(G_o/G_d,G_s)$ is the

transmission probability, i.e. the proportion of offspring with genotype G_o, when the parental types are G_d and G_s. The final term P(excluded/G_d, G_p, G_o) is one or zero, depending on whether the putative parental genotypes, G_d and G_p, could have produced the genotype G_o.

In practice, PRASE was calculated by enumerating all possible offspring genotypes for each pair of parental genotypes, then listing and checking which of the putative parental genotypes could be excluded. These were then weighted by their frequencies and summed to determine the PRASE. For example, for a single locus with two alleles *A* and *a*, if G_d and G_s is *AA* (i.e. the true parents are both *AA*), the offspring will be *AA* with a probability of 1. When G_p is *AA* or *Aa* the putative sire could be the true sire, so P(excluded/G_d, G_p, G_o) = 0. However, if G_p *aa*, the putative sire, cannot be the true sire and P(excluded/G_d, G_p, G_o) = 1. PRASE is found by summing over all possible genotype combinations.

Results

The five loci studied were found to be highly polymorphic in all the breeds of cattle studied. The *DRB3* locus had 23 alleles, ranging in size from 144–220 bp (Fig. 1a). The *CYP21* and *HEL6* microsatellite loci were also highly polymorphic in all the breeds, with 19 alleles varying in size from 254–300 bp (Fig. 1b) and 252–276 bp respectively (Fig. 1c). Twenty-four alleles with size ranging from 137–171 bp were detected for *ETH131* (Fig. 1d) while *FSHB* had 21 alleles with a size range of 182–222 bp (Fig. 1e). These markers had PIC values of 0·92, 0·91, 0·88, 0·89 and 0·91 respectively. The results for PIC and heterozygosity by breed for the different loci are given in Table 2.

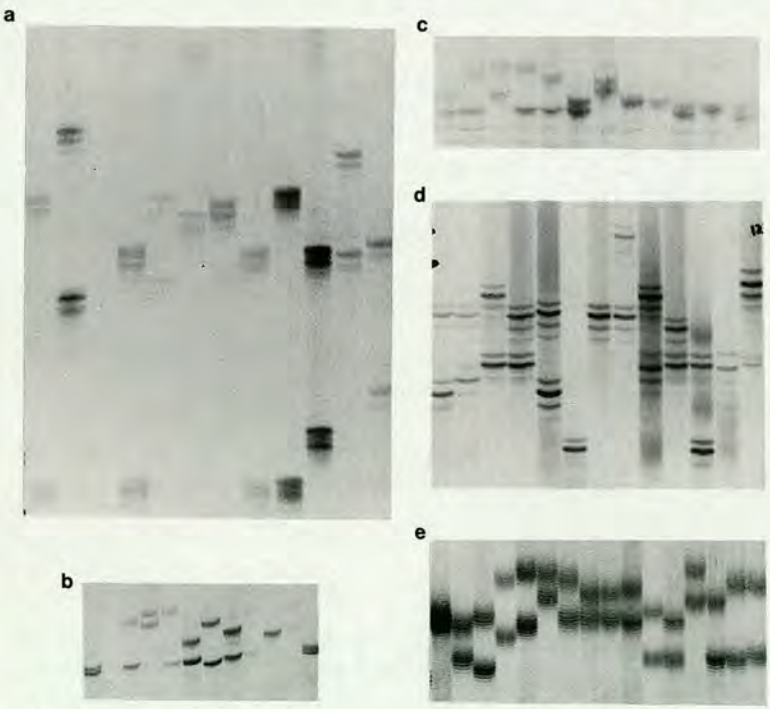


Fig. 1. Autoradiograph of denaturing 6% polyacrylamide gels with ^{35}S-labelled PCR-amplified fragments from: (a), *DRB3*; (b), *CYP21*; (c), *HEL6*; (d), *ETH131*; (e), *FSHB*.

Jersey and Simmental exhibited deviations from HW equilibrium for all the loci ($P < 0.05$) whereas Guernsey did not exhibit deviation at any of the loci ($P > 0.05$). The remaining breeds showed significant deviations from HW equilibrium at some of the loci ($P < 0.05$). Significance levels on testing for HW equilibrium are shown in Table 3.

The Monte Carlo test based on Hellinger distance, used to test for differences between breeds, compares differences in allele frequencies with the presence or absence of different alleles at the two loci. All breeds are significantly different from each other in this respect at the *DRB3* locus ($P < 0.05$). For the *CYP21* locus, similarities in allele frequencies were found between Aberdeen Angus, Guernsey, Ayrshire and Hereford, also Charolais, Jersey, Simmental and South Devon ($P > 0.4$). Belgian Blue and Guernsey were found to have similar allele frequencies for *FSHB* while Belgain Blue and Friesian were found to be similar for *ETH131* (*P* 0·2–0·4),

PRASE indicates the value of a marker for parentage verification. Calculation of PRASE depended on whether or not the loci were in HW equilibrium or linkage equilibrium. The results are shown in Table 4. PRASE values ranged from 0·44–0·85 and 0·40–0·81 for *DRB3* and *CYP21* loci respectively across all the breeds. *ETH131*, *HEL6* and *FSHB* had PRASE values ranging from 0·56–0·8, 0·14–0·81 and 0·4–0·82 respectively. The value of the markers differs between breeds, with the highest PRASE value detected in Limousin with the *DRB3* locus (0·85) and the lowest value in Belted Galloway with *HEL6* (0·14).

Taken together, *DRB3* and *CYP21* would give a 0·88 PRASE for all breeds, with success ranging from about 0·75 in the Dexter to greater than 0·96 in the Friesian. These markers are, however, closely linked (1·4 cM), and if unlinked markers with similar allele frequencies were used the overall value would be 0·91. Using all the five markers together a probability of exclusion of 99% or better can be achieved for all the 14 breeds (Table 4).

Table 2. Number of alleles, PIC and heterozygosity values for markers

Breed		Number of animals	Microsatellite loci													
			DRB3			*CYP-21*			*ETH131*			*FSHB*			*HEL6*		
			Alleles	PIC	Hetero	Alleles	PIC	Hetero	Alleles	PIC	Hetero	Alleles	PIC	Hetero	Alleles	PIC	Hetero
Aberdeen Angus	(AA)	20	12	0·86	0·88	10	0·82	0·84	8	0·76	0·4	6	0·65	0·55	9	0·73	0·55
Ayrshire	(AY)	20	9	0·83	0·85	10	0·83	0·82	9	0·75	0·75	11	0·78	0·45	12	0·84	0·5
Belgian Blue	(BB)	20	11	0·87	0·65	10	0·83	0·85	15	0·88	0·7	11	0·78	0·79	6	0·63	0·68
Blonde Aquitaine	(BA)	15	8	0·68	0·27	10	0·83	0·84	10	0·8	0·82	8	0·76	0·47	9	0·76	0·79
Charlaois	(CH)	20	11	0·87	0·88	10	0·77	0·75	7	0·75	0·78	9	0·83	0·47	5	0·68	0·53
Dexter	(DX)	15	7	0·73	0·77	6	0·69	0·73	10	0·83	0·7	6	0·68	0·72	13	0·75	0·45
Friesian	(FR)	20	11	0·87	0·45	13	0·89	0·75	12	0·79	0·81	15	0·9	0·91	11	0·83	0·65
Guernsey	(GU)	20	11	0·82	0·84	9	0·82	0·84	10	0·73	0·77	13	0·68	0·7	10	0·8	0·82
Hereford	(HF)	20	9	0·62	0·62	10	0·82	0·7	11	0·88	0·89	10	0·75	0·77	11	0·88	0·6
Jersey	(JE)	20	11	0·85	0·5	10	0·85	0·7	6	0·72	0·65	11	0·87	0·45	7	0·65	0·65
Limousin	(LN)	20	14	0·9	0·6	12	0·76	0·77	10	0·85	0·7	11	0·82	0·6	10	0·75	0·78
Simmental	(SG)	20	11	0·75	0·5	10	0·86	0·7	10	0·77	0·8	11	0·86	0·5	8	0·78	0·65
B. Galloway	(BG)	15	10	0·86	0·87	7	0·76	0·8	8	0·79	0·93	4	0·51	0·33	2	0·16	0·18
South Devon	(SD)	15	7	0·77	0·8	10	0·85	0·8	6	0·75	0·78	7	0·76	1	7	0·66	0·33

Discussion

The objective of this study was to compare the efficiency of five polymorphic, microsatellite markers in order to identify paternity errors in different breeds of cattle. Using a total of 260 animals in 14 breeds we found 23 alleles for *DRB3* and 19 for *CYP21*. In a study involving three breeds of cattle, 14 alleles were described for *DRB3* (Ellegren *et al.* 1993), while 13 alleles for *CYP21* were found in the international genome mapping reference panel (Fries *et al.* 1990). Eleven alleles were detected for *FSHB* in the CSIRO cattle reference families (Moore *et al.* 1992). In the four breeds studied, eight alleles were seen for *ETH131* (Steffan *et al.* 1993) while 10 alleles were described for *HEL6* in Finnish cattle (Kaukinen & Varvio 1993). The greater number of alleles seen for all the loci in this study may be attributed to the use of unrelated individuals from 14 breeds giving a broader genetic base than that used in the previous studies. Clearly all these markers have high levels of polymorphism, which will make them valuable for parentage verficiation in most populations.

The frequency of alleles at all loci varied considerably between breeds. Alleles present at high frequency in one breed may be absent in another (e.g. an allele of the *DRB3* locus with a frequency of 0·6 in Hereford was absent in Jersey and Friesian). There were a large number of *CYP21* alelles present in the Friesian and Limousin, covering a large size range (254–300), while fewer alleles were present in Dexter and Belted Galloway, covering a small range of sizes (260–270). With *HEL6* twelve alleles were present in Ayrshire whilst only two were found in Belted Galloway. This restricted range of allele sizes may reflect the expansion of a breed following a genetic bottleneck. The limited number of alleles seen in Dexter, which until recently was considered a rare breed, could be a result of either a small number of sires being available or of inbreeding. Comparing distribution of alleles between breeds, there was no consistent similarity for *DRB3*. However, there were some similarities with the allele distribution for other markers, probably owing to chance. The breeds that showed greatest similarity for *CYP21* alleles, a Channel Island breed (Guernsey) and Aberdeen Angus, are different in type (dairy *vs* beef) and geographical origin. Interestingly, MacHugh *et al.* (1994) have reported similarity in microsatellite allele frequencies between another Channel Island breed (the Jersey) and Aberdeen Angus.

Only a few breeds were found to be in HW equilibrium for the loci studied. In these breeds, selection, either natural or for production traits, appears not to have distorted the population. For most breeds the use of artificial insemination has enabled extensive selection programmes to be carried out, and the selection of an extreme phenotype might be expected to have produced disequilibrium. It is therefore not surprising to find that for highly selected breeds, e.g. Jersey and Simmental, none of the loci is in equilibrium. A restricted number of sires is used in these breeds. Breeds where one of the loci is in disequilibrium may reflect selection for a phenotype affected by that gene, e.g. the major histocompatibility complex is involved in antigen presentation and may influence immunity to particular diseases, whilst *CYP21* is involved in steroid metabolism and may affect production traits.

Table 3. Significance levels on testing for Hardy–Weinberg equilibrium

Breed	*DRB3*	*CYP21*	*HEL6*	*ETH131*	*FSHB*
AA	0·101	0·076	0·001	0·000	0·014
AY	0·059	0·309	0·000	0·046	0·001
BB	0·001	0·083	0·074	0·009	0·449
BA	0·000	0·114	0·148	0·331	0·000
DX	0·345	0·177	0·000	0·488	0·000
CH	0·572	0·010	0·001	0·001	0·253
FR	0·000	0·004	0·013	0·146	0·162
GU	0·734	0·050	0·053	0·322	0·445
HF	0·173	0·019	0·000	0·052	0·368
JE	0·000	0·027	0·039	0·000	0·000
LN	0·004	0·739	0·055	0·039	0·036
SG	0·000	0·002	0·004	0·004	0·000
BG	0·459	0·024	0·105	0·039	0·013
SD	0·175	0·005	0·001	0·432	0·642

Significance levels less than 0·05 indicate that the alleles may not be in Hardy–Weinberg equilibrium.

Table 4. Calculation of probability of random sire excluded (PRASE)

Breed	*DRB3*	*CYP21*	*ETH131*	*HEL6*	*FSHB*	Total
AA	0·773	0·688	0·711	0·660	0·526	0·996
AY	0·751	0·671	0·585	0·788	0·718	0·997
BB	0·820	0·730	0·804	0·416	0·595	0·997
BA	0·668	0·685	0·649	0·641	0·700	0·996
CH	0·774	0·621	0·715	0·695	0·490	0·996
DX	0·611	0·409	0·574	0·595	0·775	0·992
FR	0·831	0·812	0·661	0·753	0·810	0·999
GU	0·679	0·722	0·561	0·696	0·509	0·993
HF	0·441	0·748	0·775	0·814	0·628	0·997
JE	0·797	0·756	0·609	0·475	0·829	0·998
LN	0·852	0·586	0·765	0·612	0·730	0·998
SG	0·695	0·775	0·626	0·694	0·808	0·998
BG	0·720	0·582	0·604	0·146	0·405	0·978
SD	0·579	0·752	0·562	0·614	0·536	0·992

PRASE is a direct measure of the usefulness of a marker for parentage verification, whereas PIC and heterozygosity are only indirect measures. Previously used methods for calculating probability of exclusion assumed a single locus in HW equilibrium. Goodness of fit methods could not be used to test our data as the sample sizes were small compared with the number of alleles. We therefore used Monte Carlo, or Bootstrapping, techniques. As deviations from HW equilibrium were found, these deviations were accounted for when determining the PRASE by using observed genotypes, rather than allele frequencies. Using genotypes also allowed linkage disequilibrium to be included in the calculation of PRASE and ensured that the results were as accurate as possible, reflecting the genetic structure of the populations under investigation.

As the *DRB3* and *CYP21* loci are separated by only 1·4 cM (W. Barendse, personal communication), it is likely that they are in linkage disequilibrium, which would reduce the value of the loci if both were used. However, our results show the PRASE is not appreciably reduced, when compared with unlinked markers: for *DRB3* and *CYP21* together a PRASE of 0·89 was found for Aberdeen Angus whilst similar unlinked markers would give a value of 0·93. It should also be noted that allele distribution for *DRB3* is breed specific, and there are a greater number of alleles shared between breeds for *CYP21*. With over 300 microsatellite markers now available for cattle the use of linked markers could be avoided, however, the use of tightly linked markers may have an added advantage in that gametic association creates identifiable haplotypes that could help to identify related individuals. Assuming that all allelic combinations are possible, the *CYP21* and *DRB3* haplotypes could be considered as a single locus with 437 alleles (23 alleles × 19 alleles). To use this information, linkage phase must be determined by reference to ancestors or siblings. We are currently addressing the use of haplotypes to confirm relationships.

In conclusion, we find that two markers give 80–90% probability of excluding an incorrect sire. Including a third locus with a similar level of polymorphism increases this to about 97%, which is comparable to the estimated efficiency of parentage verification using serological techniques at the Edinburgh Cattle Blood Typing Service (R. L. Spooner & D. S. Ross, personal communication). The use of all five markers would achieve at least 99% probability of exclusion in all the breeds studied. Typing of five such microsatellite loci would be potentially achieved on one track of an electrophoretic gel, by multiplexing.

Acknowledgements

We thank Dr D. S. Ross and Ms O. Newlands of the CBTS for providing the samples. APU is funded by the Association of Commonwealth Universities, UK. This work was supported in part by MAFF OC9024.

References

Barendse W., Armitage S.M., Kossarek L.M. *et al.* (1994) A genetic linkage map of the bovine genome. *Nature Genetics* **6,** 227–35.

Bishop M.D., Kappes S.M., Keele J.W. *et al.* (1994) A genetic linkage map for cattle. *Genetics* **136,** 619–39.

Botstein D., White R.L., Skolnick M. & Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32,** 314–31.

Chakroborty R., Meagher T.R. & Smouse P.E. (1988) Parentage analysis with genetic markers in natural populations. 1. The expected proportion of offspring with unambiguous paternity. *Genetics* **118,** 527–36.

Chung B., Matheson K.J. & Miller W.L. (1986) Structure of a bovine gene for P-450c21 (steroid 21 hydroxylase) defines a novel cytochrome P-450 gene family. *Proceedings of National Academy of Sciences of the USA* **83,** 4243–7.

Creighton P., Eggan A., Fries R. *et al.* (1992) Mapping of bovine markers CYP21, PRL and BoLA DRB by genetic linkage analysis in reference pedigrees. *Genomics* **14,** 526–8.

Devlin B., Risch N. & Roeder K. (1990) No excess of homozygosity at loci used for DNA fingerprinting. *Science* **249,** 1416–20.

Devlin B. & Risch N. (1992) Ethnic differentiation at VNTR loci, with special reference to forensic applications. *American Journal of Human Genetics* **51,** 534–48.

Ellegren H., Davis C.J. & Andersson L. (1993) Strong association between polymorphisms in and intronic microsatellite and in the coding sequence of the *BoLA-DRB3* gene implications for microsatellite stability and PCR-based *DRB3* typing. *Animal Genetics* **24,** 269–75.

Fries R., Hediger R. & Stratzinger G. (1986) Tentative chromosomal localization of the bovine major histocompatibility complex by *in situ* hybridization. *Animal Genetics* **17,** 287–94.

Fries R., Eggan A. & Stratzinger G. (1990) The bovine genome contains polymorphic microsatellites. *Genomics* **8,** 403–6.

Guo S.W. & Thompson E.A. (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* **48,** 361–72.

Hetzel J. (1993) Livestock genome research on the march. *Nature Genetics* **4,** 327–8.

Jeffreys A.J., Wilson V. & Thein S.L. (1985) Hypervariable minisatellite regions in the human DNA. *Nature* 314, 67–73.

Jeffreys A.J., Macleod A., Tamaki K., Neil D.L. & Mossekton D.G. (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354,** 204–9.

Kaukinen J. & Varvio S. (1993) Eight polymorphic bovine microsatellites. *Animal Genetics* **24,** 148.

Kim K.E., Gordon D.F. & Maurer R.A. (1988) Nucleotide sequence of the bovine gene for follicle-stimulating hormone beta-subunit. *DNA* **7,** 227–33.

Litt M. & Luty J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within cardiac muscle actin gene. *American Journal of Human Genetics* **44,** 397–401.

MacHugh D.P., Loftus R.T., Bradley D.G., Sharp P.M. & Cunningham E.P. (1994) Microsatellite DNA variation within and among European cattle breeds. *Proceedings of the Royal Society of London Sec. B* **256,** 25–31.

Mejdell C.M., Lingaas F., Stear M.J. & Lie O. (1993) Genetic analysis of major histocompatibility complex class I antigens, serum transferrins and red blood cell antigens in Norwegian breeds of cattle. *Acta Agriculturae Scandinavica* **43,** 193–200.

Moore S.S., Barendse W., Berger K.T., Armatage S.M. & Hetzel D.J.S. (1992) Bovine and ovine DNA microsatellites from the EMBL and GenBank databases. *Animal Genetics* **23,** 463–7.

Muggli-Cockett N.E. & Stone R.T. (1988) Identification of genetic variation in the bovine major histocompatibility complex DR-like genes using sequenced bovine genomic probes. *Animal Genetics* **19,** 213–25.

Oliver R.A., Brown P., Spooner R.L., Joosten I. & Williams J.L. (1989). The analysis of antigen and DNA polymorphisms within the bovine major histocompatibility complex 1: the class I antigens. *Animal Genetics* **20,** 31–41.

Ott J. (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *American Journal of Human Genetics* **51,** 283–90.

Saiki R.K., Scharf S., Faloona F. *et al.* (1985) Enzymatic amplification of beta globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* **230,** 1350–4.

Saiki R.K., Gelfand D.H., Stoffel S. *et al.* (1988) Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase. *Science* **239,** 487–91.

Skow L.C., Womack J.E., Petresh J.M. & Miller W.L. (1988) Synteny mapping of the genes for 21-steroid hydroxylase, alpha-a crystallin and class I bovine leucocyte antigen in cattle. *DNA* **7,** 1453–9.

Steffan P., Eggan A., Dietz A.B., Womac J.E., Stranzinger G. & Fries R. (1993) Isolation and mapping of polymorphic microsatellites in cattle. *Animal Genetics* **24,** 121–4.

Weber J.L. & May P.E. (1989) Abundant class of DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44,** 388–96.

EXCLUDING MUTATIONS IN THE TRANSMEMBRANE REGION OF FGFR3 AS THE CAUSE OF DOMINANTLY INHERITED DEXTER DWARFISM (ACHONDROPLASIA) IN CATTLE.

Usha. A.P., Lester. D.H., Russell. G.C. and Williams. J.L

(To be presented at the Mammalian Genetics and Development Workshop, London. 20-22nd Nov 1995.)

The short-legged, dimunitive heterozygote Dexter cattle are thought to be the bovine equivalent of dominantly inherited achondroplasia (ACH/+) dwarfism in man. Homozygous Dexter cattle (ACH/ACH) or so called "bull-dog" calves, are often aborted before the eighth month of gestation, have very short legs and a distinctive craniofacial defect which causes its tongue to grotesquely protrude. Achondroplasia in humans has recently been shown to be caused by either one of two single amino acid changes in the transmembrane (TM) region of FGFR3. Our laboratory therefore isolated and sequenced most of the FGFR3 TM region in normal (+/+), heterozygous (ACH/+) and "bull-dog" (ACH/ACH) cattle. This was acheived by using human oligonucleotide primers, specific for the TM domain of FGFR3, to PCR amplify and sequence cDNA from an activated bovine (+/+) macrophage cell line. The normal bovine DNA sequence showed 89%, 85% and 63% homology with human, mouse and chicken FGFR3 respectively. Genomic DNA from Dexter carrier (ACH/+), and "bull-dog" (ACH/ACH) was then PCR amplified and sequenced using human and bovine specific oligonucleotides for the TM region. No DNA sequence difference was observed between +/+, ACH/+ and ACH/ACH cattle in the TM region of FGFR3. Moreover the two amino acids that are changed in humans to give an achondroplastic phenotype were shown to be both conserved and unchanged in ACH/+ and ACH/ACH Dexter cattle.